

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau**BL**

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 15/12, 15/54, 9/12 // C12Q 1/68, C12N 15/11	A1	(11) International Publication Number: WO 93/15201 (43) International Publication Date: 5 August 1993 (05.08.93)
(21) International Application Number: PCT/US93/00586 (22) International Filing Date: 22 January 1993 (22.01.93) (30) Priority data: 826,935 22 January 1992 (22.01.92) US (60) Parent Application or Grant (63) Related by Continuation US 826,935 (CIP) Filed on 22 January 1992 (22.01.92) (71) Applicant (for all designated States except US): NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).	(72) Inventors; and (75) Inventors/Applicants (for US only) : AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US). (74) Agents: GRANAHAH, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US). (81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: NOVEL PROTEIN TYROSINE KINASES (57) Abstract The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and
5 differentiation is regulated in part by phosphorylation of
various cellular proteins. Protein tyrosine kinases are
enzymes that catalyze this process. Moreover, many act as
growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein
tyrosine kinase genes present in human megakaryocytic and
lymphocytic cells, the proteins encoded by these genes,
antibodies specific for the encoded proteins, RNA nucleic
acid sequences which hybridize to the genes and methods of
15 use therefor.

The genes isolated as described herein are referred
to, collectively, as protein tyrosine kinase (pTK) genes.
The nucleic acid sequences of these genes, isolated as
discussed herein, show significant homology with
20 previously identified protein tyrosine kinases containing
extracellular domains which function as growth factor
receptors. The pTK genes have been shown to be present in
both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show
25 significant sequence homology with members of the c-kit
subgroup of growth factor receptors with protein tyrosine
kinase activity. The c-kit subgroup of receptor tyrosine
kinases catalyze the phosphorylation of exogenous
substrates, as well as tyrosine residues within their own
30 polypeptide chains. (Ullrich, A. and Schlessinger, J.,
Cell, 61:203 (1990)). Members of the c-kit subgroup
include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

10 The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known
15 protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120
20 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells,
25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

30 Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKs)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target
5 proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or
10 bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect
15 cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in
20 extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1
25 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

10 Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid
15 sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.
20

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid
25 sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

30 Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.

-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases
(Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and
the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30
5 respectively, also exhibit sequence homology with known
protein tyrosine kinases.

Thus, as described above, DNA which hybridize with
DNA encoding amino acid sequences present in the catalytic
domain of a protein tyrosine kinase of the c-kit subgroup
10 of protein kinases have been isolated and sequenced.
These isolated DNA sequences, collectively referred to as
pTKs genes, (and their deduced amino acid sequences) have
been shown to exhibit significant sequence homology with
known members of receptor tyrosine kinase families.

15 Once isolated, these DNA fragments can be amplified
using known standard techniques such as PCR. These
amplified fragments can then be cloned into appropriate
cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning
20 vectors, labeled with a radiolabeled nucleotide such as
³²P and used to screen appropriate cDNA libraries to
obtain the full-length cDNA clone.

The pTk genes as described above have been isolated
from the source in which they occur naturally, i.e.
25 megakaryocyte and lymphocytic cells. The present invention
is intended to include pTk genes produced using genetic
engineering techniques, such as recombinant technology, as
well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTK genes
30 include amino acid sequences which encode peptides
exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase
30 activity, or activate other growth factors.

-13-

pTk genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase
5 activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of
10 treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the
15 following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the pTK Genes

To facilitate the isolation and identification of
20 these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2). These sequences were used as polymerase
25 chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the
30 catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able
20 to ascertain using no more than routine experimentation,
many equivalents to the specific embodiments of the
invention described herein. Such equivalents are intended
to be encompassed by the following claims.

-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences
10 consisting of:
 - a) SAL-S1 (SEQ ID NOS:5 and 7);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NOS:9 and 19);
 - d) LpTK 3 (SEQ ID NO:11);
 - 15 e) LpTK 4 (SEQ ID NOS:13 and 21);
 - f) LpTK 13 (SEQ ID NO:15);
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid
20 sequence selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NOS:6 and 18);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NOS:10 and 20);
 - d) LpTK 3 (SEQ ID NO:12);
 - 25 e) LpTK 4 (SEQ ID NO:14);
 - f) LpTK 13 (SEQ ID NO:16);
 - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - 5 l) bpTK 5 (SEQ ID NO:29); and
 - m) bpTK 7 (SEQ ID NO:30).
4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain
- 10 of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:
- 15 a) SAL-S1 (SEQ ID NO:6);
 - b) SAL-D4 (SEQ ID NO:8);
 - c) LpTK 2 (SEQ ID NO:10);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTK 4 (SEQ ID NO:14); and
 - f) LpTK 13 (SEQ ID NO:16).
 - 20 g) HpTK 5 (SEQ ID NO:24);
 - h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - 25 l) bpTK 5 (SEQ ID NO:29); and
 - m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
 - a) SAL-S1 (SEQ ID NO:6);
 - b) SAL-D4 (SEQ ID NO:8);
 - 10 c) LpTK 2 (SEQ ID NO:10);
 - d) LpTK 3 (SEQ ID NO:12);
 - e) LpTK 4 (SEQ ID NO:14); and
 - f) LpTK 13 (SEQ ID NO:16).
 - g) HpTK 5 (SEQ ID NO:24);
 - 15 h) bpTK 1 (SEQ ID NO:25);
 - i) bpTK 2 (SEQ ID NO:26);
 - j) bpTK 3 (SEQ ID NO:27);
 - k) bpTK 4 (SEQ ID NO:28);
 - l) bpTK 5 (SEQ ID NO:29); and
 - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-18-

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim 10 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim 10 encoded by the amino acid sequence (SEQ ID NO:8).
16. A protein of human megakaryocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
- 15 17. A protein of Claim 14 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim 14 encoded by the amino acid sequence (SEQ ID NO:12).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:5);
 - b) SAL-D4 (SEQ ID NO:7);
 - c) LpTK 2 (SEQ ID NO:9);
 - d) LpTK 3 (SEQ ID NO:11);
 - e) LpTk 4 (SEQ ID NO:13); and
 - f) LpTK 13 (SEQ ID NO:15).
 - g) LpTK 25 (SEQ ID NO: 22 and
 - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

1/35

sals1 (160 bases)
FLKI-LIKE

1	PTKI/3 PRIMERS	21	41
5' <u>ggatcctgtgcatcagtgactgacttagggctagggaacattctgctgctggaagcgacgtggt</u>			
		D P V H Q a L R A R N I L L S E S D V V	
61		81	101
gaagatctgtgactttggccttggccctggacatctacaagagacccagctacgtccgcaa			
		K I C D F G L A R D I Y K D P S (Y) V R K	
121		141	PTKKW PRIMER
gcattgccggctgccctgaagtggatggcgccagaattc 3'			
		H A R L P L K W H A P E F	

FIGURE 1

sald4 (147 bases)-
FGFR-LIKE

1	<u>PTKI/3 PRIMERS</u>	21	41
6'	<u>ggatccattcacagagacctagcagcagcaacatcctggtctctcagaggacctggtaac</u>		
	G S I H R D L A A R N I L V S E D L V T		
61	<u>81</u>	101	
	<u>aagtcagcgactttggcctggccaaagccgagcggaagggtctagactcaagccggctg</u>		
	K V S D F G L A K A E R K G L D S S R L		
121	<u>PTKKW PRIMER</u>	141	
	<u>ccggtcaaatggatggctcccgaattc</u> 3'		
	P V K W H A P E F		

FIGURE 2

3/35

LpTK2

GTTGGAATTCCTTCCGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG
ATATTATGTTACCAACGAGGACATTCCT

FIGURE 3A**LpTK3**

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAACACCCTCTCGAAAGTT
GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

FIGURE 3B**LpTK4**

G TTCACCGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATTGAGAGTGACGAC
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC
ACACAAATGAGTGCCGC

FIGURE 3C**LpTK13**

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC
AGTGATTTCCGACTTTCCAAAGCACTGCGTGCTGATGAAACTACTACAAGGCCAGACC
CATGGAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC
AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

FIGURE 3D

1 TTGAGCTCG CCGACNTTG ATTATGACT AGTATTATAT TACGGGCTCA TTACTTCATA GCCATATAT GGAGTTCGGC GTTACATAAC
AAGCTCGAGC GGGCTGTAC TAATAACTGA TCATAATTA TCNTTAGTTA ATGCCCCAGT ATCAAGTAT CGGGTATATA CCTCANGGCG CAATGTATTG

101 TTACGGTAA TGSCCGCTT GGTGACGCG CCAACGACC CCGCCCATTTG ACGTCANZAA TGACGTATGT TCCCATAGTA ACCCAATAG GGACTTTCCA
AATGCCATTT ACCGGGCGGA CCGACTGCGG GGTTCCTGGG GCGGGGTAC TGCAGTATTT ACTGCATACA AGGTATCAT TCGGTATATC CCTGANAAGGT

201 TTGACGTCAA TGGGTGGAGT ATTTACGGA AACTGACCAC TTGGCAGTAC ATCAAGTGA TCATNTGCCA AGTACGCCC CTATTCACGT CAATGACGGT
AACTGCAGT ACCCACTCA TAAATGCCAT TTGACGGGTG AACGTCATG TACTTCACAT AGTATACGT TCATGCGGG GATAACTGCA GTTACTGCCA

301 AAATGGCCCG CCTGGCATTA TGCCACGTAC ATGACCTAT GGCACCTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGTATGC
TTATCCGGGC GGACCGTAT ACGGGTCATG TACTGGNATA CCTGAAAGG ATGAACCTC ATGTAGATSC ATAATCAGTA GCGATATAGG TACCACTAGC

401 GGTTTTGGCA GTACATCAT TGGCGTGGAT AGCGTTTGA CTCACCGGA TTCCACCTC TCCACCCCAT TGACGTCAAT GCGAGTTTGT TTTGGCACCA
CCAAACCCGT CATGTAGTTA CCGGCACCTA TCGCCAAACT GACTGCCCTT AAGGTTTAC AGGTGGGTA ACTGCAGTA CCTCAACA NAACCGTGGT

501 AAATCAACGG GACTTTCCA ATGTGCTAA CAACTCCGC CCATGACCG CCATGGCGG TAAGGCTGA CGTGGGAGG TCTATATAAG CAGAGCTCGT
TTTAGTTGCC CTGAAGGTT TTACAGCAT GTTGAGGCGG GTTAACTCG GTTACCGGC ATCCGCACAT GCCACCCCTC AGATATATTC GTCTCGAGCA

601 TTAGTGAACC GTCAATCCG CTGGAGACCG CATCACCGCT GTTTGACCT CCATAGAGA CACCGGAGC GATCCAGCT CCGCGGCTCG GAACGGTGCA
AATCACTTGG CAGCTAGCG GACCTCTCGG GTAGTGCGA CAAACTGGA GGTATCTTCT GTGGCCCTGG CTAGGTGGA GCGCGCGGCC CTTCGCCACGT

FIGURE 4A

5/35

701 TTGCAACGGG GATTCCCCCT GCCAAGACTG ACPPACTAC CGCTATAGA GTCTATAGG CCAATTGGCT TCGTTAGAAC GCGCTACAA TTAATACATA
AACCTTGCG CTAAGGGGA CGGTTCTCAC TGCATTCATG GCGATATCT CAGATATCG GGTGNAACCA AGCANCTTG CCGCGATGT ANTATGTAT

801 ACCTTATGTA TCATACACAT ACGATTAGG TGACACTATA GAATAACATC CACTTGCCT TTCTCTCCAC AGGCTCCAC TCCAGGTCC AACTGCACCT
TGGAAATACAT AGTATGTGA TGTAAATCC ACTGTGAT TTTATTTAG GTGAACCGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGA

901 CGGTTCTATC GATGAATTC CCGGGGATC CTCATAGAGT CCCTCGACCT CGAGATCCAT TGTGCTGGC CGGATTCCTT ATCACTGATA AGTTGGTGA
GCCAAGATAG CTAACCTAAG GGGCCCTAG GAGATCTTA GAGATCTTA GGTCTAGGTA ACACGACCCG GCTAAGAA TACTGACTAT TCAACCACT

1001 CATATTATGT TTATCACTGA TAAAGTGTA ACCATGACAA AGTGCAGCC GAATACAGTG ATCCGTGGC CCTAGACCT GTTGAAACGAG GTGCGGCTAG
GTATAATACA AATAGTCACT ATTACACAGT TCTACTGT TCAACGTCCG CTTATGTCAC TAGGCACGGC GGCATCTGA CACTTCTC CAGCCGATC

1101 ACGGTCTGAC GACAGGCMAA CTGGCGGAC GGTGGGGGT TCAGACGCC GCGTTTACT GGCATTTCAG GAACAAAGCG GCGCTGCTCG ACGCACTGGC
TGCCAGACTG CTGTGCGTTT GACCGCCTTG CCAACCCCA AGTCCTGCG CCGCAATGA CCTGAAGTC CTGTTCGCC CCGGACGAGC TGGGTGACCG

1201 CGAAGCCATG CTGGGGGAGA ATCATAGCAC TTGGGTGCC AGAGCCGACG ACGACTGGC CTCATTTCTG ACTGGGATG CCGCAGCTT CAGGAGGGG
GCTTCGGTAC GACCGCTCT TAGTATCTG AAGCCACGGC TCTGCGCTGC TGTGACCGC GAGTAAAGAC TGACCTTAC GGGCTCGAA GTCCGTCCG

1301 CTCCTGCTC ACCCCAGCA CAATGGATCT CGAGGATCT TCCATACCTA CCAGTTCTGC GCGTGCAGGT CGCGGCCGA CTACTCTTG ATGTATTACT
GACGAGCGGA TGGCGGTCGT GTACCTAGA AGGTATGAT GTTCAGACG CCGACGTCCA GCGCCGGCT GATGAGAAC TACATATGA

FIGURE 4B

6/35

1401 CATATTACCA AGGAATAACT GCGGGGCACA GGGTCAGGTG CTGAGGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT GTACACGCTGG
 GTATATGGT TCCTATTGA CCGCCCGTGT CCGAGTCCAC GACTTCCTG TACACACTT CACTGGATCT TCGGTCTCC ACTCGGAGA CAGTGGCACC

1501 CATANGGCC GCTGAGGGC TCTTGTGTA AGCACTAACG CCACTGTCTG GGAGGCCACC TGTACTCTAG CAGACCAATGA AAGGCGTCT CCTTTTCCT
 GTATCCCGG CCACTCCG AGAACCCAGT TCGTCAATGC GGTACACAGC CCTCCGTGG ACAATGACT GTCTGTAAT TTCCCGCAGA GGAAGAGAA

1601 GCAGGAGTCA GGAACACTC TGCTCCACCA GCTTCTTGTG GGAGGCTGGA TATTATCCAG GCTGGCCCG AGTCATCCG AGCCTAAC CCTCCCTGTG
 CCTCCTCAGT CCTTGTGAG ACGAGGTGCT CGAAGAACAC CCTCCGACCT ATANTAGGTC CGGACGGGCG TCAGTAGGCC TCCGATTTGG GGAGGGACAC

1701 GTGCTTCACT GGTACACTC CTGTGCACT TCTATGCTCC TCTTGGCTC TGCTGCTTC TTGTAAGTT GTATAGATA GCAGAGAAA TAGCGAAGT
 CACGAAGTCA CCAGTGTGAG GAAACAGTGA AACTACGAGG AAGACGGGAG GACCAAGGAG AACCTTCAAA CATCACTAT CGTCTCTTT ATCGCTTTCA

1801 CTTAAGTCT TTGATCTTC TTATAAGTGC AGAGNAGAAA TGCTGACGTA TGCTGCTTC TCTCTCTG TGCTGCTAC CTGAAGCCGC TTTCTTGTCT
 GAATTCAGA AACTAGAAAG AATATTACG TCTCTCTTT ACGACTGCAT ACGACGGGAG AGAGAGAGAC GAAATCGATG GACTTCGGG AAGAGAACAGA
 349 O R P G S E O R

1901 ATACCTGCTC TCTATCTGCT CACACTCCTC CGAGGCCAGC ACCATCCAC TGTCTGTCTG GTTGTCCACA GAGCCTTCT AGGTGCTTGG GGTCTATCGG
 TATGGAGAG AGATAGACGA GTGTGAGGAG GCTCCGGTCT TGGTAGGCTG ACAGACACAC CAACAGGTGT CTCGNAACA TCCAGCAACC CCAGTACCC
 341 Y R S E I O E C E E S A L V H G S D T Q N D V S G K Y T T P T H P

2001 ATTCTTCAA ATGTCTTCAT CTTGGAGGAA CCACGGGTCT CAGCCCTCT GGCCAGGCNC CCGGAAGG ACACCCAGT GTANTACTG GCGGCCAGGC
 TTAGGAGTT TACAGAATA GACCTCTCT GTGCCCGAGA GTCCGGGAGA CCGTCCGTG GGCCTTTTC TGTGGGTCAA CATATGGAC CGCCGTCGG
 308 P E E F T K H R S S G R T E A C R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGGCTG CAGGCTTGGC GGGTGTCTT CAGGCTCAGC CTGGCGATG TGTAGGGCCA TGGTGGACAC CTGGGAGAG CTGCTCTTT CTGAGCTGTG
 ACACCGGAC GTCCGACCG CCCGACAGG GTCCGATCG GACCGCTAC ACATCCGGT ACCACCTGG GACGCTCTT CAGCGAGAA E S S Q
 215 S H R Q L S P P S D E A D A Q A I H L A M T S V Q S F S G E E S S Q

2201 AGAGCTGGC GGGGCCATCC AGACTCTTC TTCTCTTGC AGGCCCCCTG CCTGGAGCAG GTCCCCGAGG ATCTCCACCA GCTCCGAGAA TGAAGGTCTC
 TCTCGACCG CCCCGTACG TCTGGAGGAG AAGGAGACC TCCGGGACG GACCTCTC CAGGGCTCC TAGAGTGT CGAGCTCTT ACCTCCAGC
 241 S S R P A M C V E E E Q L G R G O L L D G L I E V L E S F A P R

2301 GCCTTGGGT CTCCGACCA GCAGTTCAGC ATGATGGCG GTATGGCGG AGTGGCCAGC TCCGGGGCCC TCATCTTGT CCCGCTCTC AGCCGCTGGC
 CGGAACCCA GAGCGTGT CCGAAGTCG TACTAGCG CATACGCC R I A P T A L E P A R M R T G D R L R Q
 208 A K P D G S W C N L M I R R I A P T A L E P A R M R T G D R L R Q

2401 AGACTCTCTC ATGATCTGC AGCCAGGGT ACGGGAGGC CCCAGAGAG AAGTCTCC AGAGAGCAC CCCAAGGAC CACAGCTAC TCTGGTGTG
 TCTGAGG TACTAGAG TGGGTCCCA TCCCTCTCC GGGTCTCTC TTCTAGAGG TCTCTCTG GGTTCCTG CTCTCAGT AGACGCACCA
 175 C F E E N I Q V G P Y P S A G L S F I E W L L V G P S W V D S Q T T

2501 GTACACTTG TCGAGATGC TTTCAGGGC CATCCACTTC AGGGCAGCC GGGCACTGC CTGGCGAGC TAGTGGGGT CTTCGTAGT GTCCCGGCA
 CATGTGGAC AGCTTCTAG AAGTCCCG GTAGTCAAG TCCCGTCCG CCCGTGACG GAACGCTGC ATCAGGCCA GAACATCTA CAGGCTCTG
 141 Y V K D F Y S E P A M W K L P L R A S C K R V Y D P D K Y I D R A

2601 AGGCCAAGT CACAGATCTT CACCACGTC CTTCGACCA GCAGATCTT CCGAGCAGCC AGGTCTCTT GGATGACTT TCGGAGCC AGGAATCCA
 TCCGCTTCA GTGTAGAA GTGGTCCAC GAAAGGCTT CGTCTTCAA GGTCTCTGG TCCAGACCA CCTAGCTGA AGCCTCTCG TCTTGAGT
 108 L G F D C I K V V D S E S L L I N R A A L D R H I C K R S A L P E

2701 TCCCTCTGC CAGCTGGAG CTGTAGCAGA CAAGATCTT CATGTTCAGC GGGCTCAGC ACAGTCTCT AGCTTCTTG TCTGGAGAG CCCGCTCTG
 AGGAGACCG GTGACCTTC GACATCTCT GTCTAGAG GTACCATCG M T L P S L W L D E A E Q D P S A R R A
 75 M G R A V Q F S Y C V L D E H T L P S L W L D E A E Q D P S A R R A

FIGURE 4D

FIGURE 4E

2801' TCCGCCCTCG GTCTTCUAGA ACCGCGGCAA GAGGACCTTG TCGTGTCTCC CCGGCCCGCT CCGATCCAGC CTGGCGAGCT CCACCATGGC GCGGAAGCGT
 AGCGGGGAGC CAGAAGCTCT TGGCGGCTT CTCTGGGAC AGCGAGGAGG GGCCGGCGGA GGCTAGGTGG GACCGCTGGA GTGTGTACCG CGCTTTCGCA
 41' C G E T K S F R A F L V R D S S G P R R R D L R A L E V M A R F R
 2901' CCGCGGTGCT CGGAGACTT CTCTCGGGA TGCACGAGC TGGTCGNGG GGGCCAGTC GTCCGCCGA GAGCGGCTC CATTCGCCCG CCGCGCGGG
 GCGCGGACGA GCCCTCTGA GAGGACGCT ACCTGCTTCG ACGAGCTCC CCGGGTTCAG CAGCGGCTC GTCCGCCAG GTAGCGGC GCGCGCGGC
 8' G R Q E P S K E
 3001' CCGCCCGCAG CCGGCCCGT CACCGKBCAG GGGTGGGG CCGACTCTA GATCGACCT GAGAGGCTT GGGCGCCTG GCGAAGCTG TTTATTCAG
 GCGGGCGTC CCGCGGCGA GTGGC?CTC CCGAAGCGG GCGGTGAGT CTCAGCTGA CTTCTTCAA CCGCGGTAC CCGGTTCAG AATATCCTC
 3101' CTTATAATGG TTACAAATA AGCAATAGCA TCACAAATTT CACAATAA GCATTTTTT CACTGCATC TAGTGTCTT TTGTCCAAAC TCATCATCT
 GAAATATACC AATGTTTAT TCCTTATCT AGTGTATAA GCTTTATTT CGTAAAAA GTGACGTAAG ATCAACACCA AACAGGTTG AGTAGTTACA
 3201' ATCTATCAT GTCTGGATCG ATCGGGAATT AATCGGGCTT AATCGGGCT AGCACATGG CTTCAATAA CTTCTGAAG AGCACTTGG TTAGGTACCT TCTGAGCGG
 TAGAATAGTA CACACCTAGC TAGCCCTTAA TTAGCCCGG TCGTGGTACC GGACTTTAT GGACACTTC TCCTTGAACC AATCCATGGA AGACTCGGC
 3301' AAGAACCA GCTCTGGATG TGTGTAGTT AGGTGTGGA AATCGCCAG GTTCCCGCAG AGCAGAGGT ATGCAAGCA TGCATCTCA TTAGTCAGCA
 TTTCTTGGTC GACACCTTAC ACACAGTCA TCCACACCT TCCAGGGTC CGAGGGTCC TCCCTCTCA TACGTTCTT ACGTAGAGT AATCAGTCT
 3401' ACCAGGTGT GAAAGTCCC AGGCTCCCA GCAGGAGAA GTATGCAAG CATGCTCTC AATAGTCAG CAACCTACT CCGCCCGCTA ACTCCGCCCA
 TGTCCACAC CTTTCAGGG TCCGAGGGT CGTCCGCTT CATACGTTT CATAGTAG GTTATCAGT GTTGGTATCA GCGCGGGAT TCAGCGCGGT

FIGURE 4F

3501 TCUCGCCCCC AACTCCGCC AGTCCGCC ATTCCTCGCC CCATGGCTGA CTNATTTT TTATTATGC AGAGGCCGAG GCCGCTCGG CCTCTAGCT
 AGGGCGGGG TTGAGGCGGG TCAAGCGGGG TAAGAGCGGG GGTACCGACT GATHAANAA AATAANTACG TCTCCGCTC CGCGGAGCC GGAGACTCGA

3601 ATTCCAGAG TAGTGAGGAG GCTTTTGG AGGCTAGGC TTITGCAAAA AGCTGTAAAC AGCTGGCAC TGGCCGTGCT TTACAACTT CGTGACTGGG
 TAAGGTCTC ATCCTCTCTC CCAANARACC TCCGATCCG AAACCTTT TCGACATTTG TCGNACCGTG ACCGGCAGCA AATGTTGCA GCACCTGACCC

3701 AAAACCCCTGG CGTACCCCA CTAAATCGC TTGCAGCACA TCCCCCTTC GCACTCTGC GTANTAGCA AGAGGCCCG ACCGATCGC CTTCACAACA
 TTTTGGGACC GCANTGGGT GANTAGCG AACGTCTGT AGGGGGGAG CGCTGACCG CATATCGCT TCTCCGGCG TGGTAGCGG GAAGGTTGT

3801 GTTGGTAGC CTGAATGGG AATGGGCTT GATGGGTAT TTCTCTTA CGATCTGT CGGTATTCA CACCGATAC GTCAAGCAA CCATAGTACG
 CAACCATCG GACTACCG TTACCGGGA CTACGCCA AAAGAGAT GCTAGNAC GGTATAGT GTGGGTAT CAGTTCTGT GTATCATCG

3901 CGCCTGTAG CGGCGCTTA AGCGCGCG GTGTGTGT TACGCGCAG GTGACCGCTA CACTTGGCAG CGCCCTAGCG CCGCTCTT TCGTTTCT
 GCGGACATC GCGCGTAT TCGCGCCGCC CACACCACCA ATGCGCTCG CACTGCGAT GTGAACGGTC GCGGATCGC GGGGAGGA AGGNAAGA

4001 CCCTTCTTT CTGCCCAGT TCGCGGCTT TCCCGCTCA GCTCAATC GGGGCTCC TTATGGTTC GATTTAGT GTTACGCA CCGGACCG
 CGGAAGGAA GAGCGGTGA AGCGGCCGA AGGGGCTT CGAGATTAG CCGCGAGGG AATCCCAAG GCTAATCAC GAATGCCGT GGAGCTGGG

4101 AAAAATCTT ATTTGGTGA TGGTCACT AGTGGGCT AGCCGATA GCGGCTTT GACGTTTTT CGCCCTTCA GGTGGAGTC CACGTTCTT AATAGTGGAC
 TTTTGTGAC TAAACCCACT ACCAAGTGA TCACCGGTA GCGGCTAT CCGCAAAA CCGGGAAT GCAACCTCAG GTGCAAAA TTATCACCTG

10/35

FIGURE 4G

4201 TCTTCTTCCA AACTGGGAACA ACACCTCAACC CTATCTCGGG CTATCTCTTT GATTATTAAG GGATTTTGGC GATTTCGGCC TATTGGTTAA AAATGAGCT
AGAACAAAGT TTGACCTTGT TGTGAGTGG GATAGAGCCC GATAGAAA CTAATATTC CTAAGCCCG ATACCAATT TTTTACTCGA

4301 GATTTAACA AAATTAAACG CGAATTTTAA CAAATATTA AGTTTACAA TTTATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCCG ATAGTTAAGC
CTAAATTTGT TTTAANTGC GCTTAAATTT GTTTATATAT TGCNAATGT AANAATACCAC GTGAGAGTCA TGTATAGCCA GACTACGGCG TATCAATTCG

4401 CAACTCCCT ATCGCTACGT GACTGGGTCA TGGCTGGCC CCGACACCCG CCAACACCCG CTGACGGGT TGTCTGCTCC CGCATCCCG
GTTGAGGCGA TAGCGATCCA CTGACCCAGT ACCGACGGG GGTGTGGC GACTGCGCG GACTGCCCGA ACAGACGAGG GCCGTAGCG

4501 TTACAGACAA GCTGTGACCG TCTCCGGAG CTGCATGTGT CAGAGTTTT CACCGTCATC ACCGAAACG CGGAGGCACT ATCTTTGAAG ACCAAAGGC
AATCTCTGT CGACACTGCC AGAGGCCCTC GACGTACACA GTCTCAAA ATCTGCAAG GTGGCAGTAG TGGCTTGG CGCTCGTCA TAAGACTTC TGTCTCCCG

4601 CTCCTGATAC GCGTATTTT ATAGGTAAT GTCATGATAA TAATGTTTC TTACAGTCA GGTGCACTT TTCGGGAAA TGTGGCGGA ACCCTATTT
GAGCACTATG CGGATAAAA TATCCAAATTA CAGTACTATT ATTACCAAG ATCTGCAAG CCACCTGAA AGCCCCCTT ACACGGCCT TGGGNTAA

4701 GTTATTTT CTAATACAT TCAATATGT ATCCGCTCAT GAGCAATAA CCTGATAA TGCTCAATA ATATTGAAA AGGAGAGTA TGAATTTCA
CAATAAAA GATTATGTA AGTTATACA TAGCGAGTA CTCTGTAT GGGACTATTT ACCAAGTTAT TATACTTTT TCCTTCTCAT ACTCATPAAT

4801 ACATTTCCCT GTGGCCCTTA TTCCCTTTT TGGGGCATTT TGCCTTCTG TTTTCTCA CCGAGAACG CTGGTAAG CTGAAGATCAG
TGTAAAGCA CAGCGGGAAT AAGGCAAAA ACCCGTAA ACAGGAGGAC AAAACCGAGT GGTCTTTGC GACCACTTC ATTCTTACG ACTCTAGTC

FIGURE 4H

4901 TTGGGIGUAC GAGTGGGTTA CMTGMACTG GATCTCAACA GCGGTAAAGT CTTGAGAGAT TTTCGCCCCG AAGAACGTTT TCCANTGATG AGCATTIT
 AACCCACGTG CTCACCCCAAT GTACCTTGAC CTAGAGTTGT CCGCATTTCTA GGACTCTCA AAGCCGGGGC TTCTTGCAAA AGGTACTATC TCGTGAANA

5001 AAGTCTCTCT ATGTGGCCCG GTATTATCCC GTGATGACCG CGGGCAAGAG CAACTCGGTC GCCGCATACA CTATTCTCAG AATGACTTGG TTGAGTACTC
 TTCAGACGA TACACCCGCG CATATAGGG CACTACTGG GCCCGTCTC GTTAGCCAG GTTCTATGT GATAAGATC TTAAGTGAAC AACTCATGAG

5101 ACCAGTCACA GAAGGCATC TTACGGATGG CATGACAGTA AGAGATTAT GCAGTGTGCG CATACCAAG AGTGATAACA CTGCGGCCAA CTTACTTCTG
 TGGTCAGTGT CTTTTCGTAG ATGCTTACC CTACTGTCT TCTCTTATA CGTCACGACG GTATTGGTAC TCACATTTGT GACGCCGTT GAATGAAGAC

5201 ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTCG ACACATGCG GGATCATGTA ACTGCGCTTG ATCGTTGGGA ACGGAGCTG AATGAAGCCA
 TCTTCTAGC CTCCTGGCTT CCTCGATTGG CGAANAACG TGTCTACCC CCTAGTACAT TGACGGGAAC TAGCAACCT TGGCCTGAC TTACTTCTGT

5301 TACCAACGA CGAGCGTGAC ACCACGATGC CAGCAGCAAT GCCACAACG TTGCGCAAC TATTAACTGG CGAATCTATT ACTCTAGCTT CCGCCCAACA
 ATGGTTTGT CCTCGCACTG TGGTGCTAG GTGCTGTTA CCGTGTTC AACCGTTT ATATTGACC CTTTATGAA TGAGATGAA GGCCCTTGT

5401 ATTAATAGAC TGGATGGAG CGGATTAAGT TGCAGACCA CTCTGCGCT CGGCCCTTC GGCTGGCTGG TTATGCTG ATTAATCTG AGCCGGTGAG
 TAATATCTG ACCTACCTCC GCTATTTCA ACCTCTTCA CAGACCGCA GCCGGNAG CCGACCGACC AATAACGAC TATTAGACC TCGCCCACTC

5501 CGTGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGAG TCAGGCAACT ATGATGAAC
 GCACCCAGAG CCGCATAGTA ACCTCTGAC CCGCTCTAC CATTCGGAG GGCATAGCAT CANTAGATGT CTTGCCCTC ACTCCGTTGA TACTACTTG

FIGURE 4I

5601	GAATAGACA	GATCGTGAG	ATAGTGCCCT	CACGTATTAA	GCATTGGTAA	CTGTACAGCC	AGCTTACTC	ATATATCTT	TAGATTGATT	TAAACTTY
	CTTATCTGT	CTAGCGACTC	TATCCACGGA	GTGACTAATT	CGTAACCAT	GACAGTCTGG	TTCAAATGAG	TATATATGAA	ATCTAACTAA	ATTTTGAA
5701	TTTTTAATT	AAAAGGATCT	AGGTGAAGAT	GCTTTTTCAT	ANYCTCATGA	CCAAATCCC	TAAACGTGAG	TTTTCGTTCC	ACTAGCGTC	AGACCCCGTA
	AAAAATTAA	TTTTCTAGA	TCCACTTCTA	GGAAAAACTA	TTAGAGTACT	GCTTTAGGG	AAATGCACTC	AAAGGCAAGG	TGACTCGCAG	TCTGGGGCAT
5801	GAAGAATCA	AAGGATCTTC	TTGAGATCCT	TTTTTTCTGC	GCGTAATCTG	CTGCTTGCAA	ACAAAAAATC	CACCGCTACC	ACCGTGCTT	TGTTTGC6GG
	CTTTTCTAGT	TTCTTAGAAG	AATCTTAGGA	AAAAAAGACG	CGCATTAGAC	GACGAACGTT	TGTTTTTTG	GTGCGGATGG	TGCCCACCAA	ACAAACGGCC
5901	ATCAAGAGCT	ACCAACTCTT	TTTCCGAAGG	TAACGTGGCT	CACGAGAGCG	CAGATACCAA	ATACTGTCTT	TCTAGTGTAG	CGGTAGTTAG	GCCACCACTT
	TAGTTCTCGA	TGTTTGAGAA	AAAGGCTTCC	ATTGACCGAA	GTGCTCTCCG	GTCTATGCTT	TATGACAGGA	AGATCAGATC	GCGATCAATC	CGTGCTGAA
6001	CAAGAACTCT	GTAGCACCGC	CTACATACCT	CGCTCTGTCTA	ATCTGTGTAC	CAGTGGCTGC	TGCCAGTGGC	GATAATCTGT	GTCTTACCGG	GTGGACTCA
	CTTCTTGAGA	CATCGTGCGG	GATGTATGGA	GCGAGACGAT	TAGGACAATG	GTACCCGACG	ACGGTCACCG	CTATTACGCA	CAGANTGGCC	CAACCTGAGT
6101	AGACGNATGT	TACCGGATAA	GGCGGACGGG	TGCGCTGMA	CGGGGGGTTC	GTGCACACAG	CCCACCTTGG	ACCGACGAC	CTACACCGAA	CTGAGATACC
	TCTGCTATCA	ATGGCCCTATT	CCGCGTGGCC	AGCCCGACTT	GGCCCCCAAG	CACGTGTGTC	GGGTGGAACC	TGCTTGCTG	GATGTGGCTT	GACTCTATGG
6201	TACAGGCTGA	GCATTGAGAA	AGCGCCACGC	TTCCCGAAGG	GAGNAAGGCG	GACAGGTATC	CGGTAAGCGG	CAGGGTCCGA	ACAGGAGAGC	GCACGAGGGT
	ATGICGGACT	CGTAACCTCT	TGCGGGTGGC	AAGGGCTTCC	CTCTTTCCGC	CTGTCCATAG	GCGATTCGCC	GTCCAGCCT	TGCTCTCTCG	CGTGCTCCC

13/35

FIGURE 4J

6301 GCTTCCAGGG GGAACGCCT GGTATCTTAA TAGTCCTGTC GGGTTTCGCC ACCGTGACT TGAGCGTCGA TTTTCTGAT GTCGTCAGG GGGCGGAGC
CGAAGGTCCC CCTTTGGCGA CCATAGAAAT ATCAGGACAG CCCAAGCGG TGCAGACTGA ACTGCGAGCT AAAACACTA CGAGCAGTCC CCCCCTCTG

6401 CTATCGAAA ACGCCAGCA CCGGCGCTTT TTACGGTTCC TGGCCTTTG CTGGCCTTTT GTCACATGT TCTTCTCTC GTTATCCCTT GATTCTGTG
GATACCTTT TCGGCTCGT TCGCCGGA AATGCCAAG ACCGGAAC GACCGAAG CAGGTGTACA AGNAGGAGC CAATAGGGA CTAGACACC

6501 ATAACCTAT TACCGCCTTT GAGTGAGTG ATACCGCTCG CCGCAGCGA ACGACGAGC GACGCGAG AGTGAGGAG GAAGCGGAG ACGGCCAAT
TATTGGCATA ATGGCGGAA CTCACTCGAC TATGGCGAG GCGTGGCT TGTGGCTCG CGTGGCTCAG TCACGCGCTC CTTGCGCTT TCGCGGCTTA

6601 ACGCAACCG CCTCTCCCG CGCGTTGGLC GATTCATTAA TCCAGCTGC ACGACAGTT TCCCGACTCG AAGCGGCA GTGAGCGCA CGCAATTAAT
TCCGTTTGGC GGACAGGGG GCGCAACCG CTAAGTAAT AGGTGAGCG TCGTGTCCA AGGCTGACC TTTGCGCGT CACTCGCTT CCGTTAATTA

6701 GTCAGTACC TCACATTA GGCACCCAG GCTTACACT TTATGCTTC GGTGCTATG TTGTGTGGA TTGTGAGCG ATAACTAT CACACAGGA
CACTCATGG AGTGAGTAAT CCGTGGGTC CGAATGTGA AATACGAGG CCGAGCATAC AACACCTT AACACTGCC TATGTATA GTGTCTCTT

6801 ACAGCTATGA CCATGATTAC GAATTAA
TGTGAGACT GGTACTAATG CTAAAT

FIGURE 5A

```

1  TTGAGAGCTCG CCCGACATGTG ATATATGACAT AGTATATTAAT TACGGGTCTCA TTAGTTCATA GCCCATATAT GCCGTTCCGC GTTACATATAC
   AACCTCGAGC GGGCTGTAAAC TAAATAACTGA TCAATAAATTA TCATTAATTTA ATGCCCCAGT ATCAAGTAT CCGGTATATA CCTCAAGGCG CAATGTATTTG

101 TTACGGTAA TGGCCCGCCT GCGTGACCGC CCACGACCC CCGCCCATTTG ACGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA
   AATGCCATTT ACCGGGCGGA CCGACTGGCG GGTGTCTGG GCGGGGTAAAC TGCACTTATT ACTGCATACA AGGTATCAT TCGGTTATC CCTGAAAGGT

201 TTGACGTCAA TGGGTGGAGT ATTTACGGTA AACTGCCCAC TTGGCAGTAC ATCAAGTGT TCAATATGCCA AGTAGGCCCT CTATTGACGT CAATGACGGT
   AACTGCAGTT ACCCACCTCA TAAATGCCAT TTGACGGGTG AACCGTCA TGTTTCACAT AGTATACGGT TCATGCGGGG GATAACTGCA GTTACTGCCA

301 AAATGGCCCG CCTGGCATTA TGCCCAAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTCAATG
   TTTACCGGGC GGACCGTAAT ACGGTCATG TACTGGATA CCTGAAAGG ATGMAACGTC ATGTAGATGC ATAATCAGTA GCGATAATGG TACCACATAG

401 GGTTTGGCA GTACATCAAT GGGCGTGGAT AGCGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA
   CCMAAACCGT CATGTAGTTA CCGCACCTA TCGCCMAACT GAGTGCCCT AAAGGTTGAG AGGTGGGTA ACTGCAGTTA CCTCNAACA AAACCGTGGT

501 AAATCAACGG GACTTTCCAA AATGTCGTAA CAACTCCGCC CCATTGACGC AAATGGGCGG TAGGCTGTGA CGGTGGGAGG TCTATATAAG CAGAGCTCGT
   TTTAGTTGCC CTGAAAGGTT TTACAGCATT GTTGAGCGG GGTAACTGCG TTACCCGCC ATCCGCACAT GCCACCTCC AGATATATTG GTCTCGAGCA

601 TTAGTGAACC GTCAGATCGC CTGGAGACGC CATCCACGCT GTTTGACCT CCATAGAAGA CACCGGACG GATCCAGCCT CCGCGGCCCG GAACGGTGCA
   AATCACTTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCGA CAAACTGGA GGTATCTTCT GTGGCCCTTG CTAGGTGGA GCGCGCCGCC CTGCCCCAGT

701 TTGGAAACGG GATTCCCCGT GCCAAGAGTG ACGTANGTAC CGCTATAGA GTCTATAGGC CCACCTGGCT TCGTTAGAAC GCGGCTACAA TTAATACATA
   AACCTTGGCG CTAAAGGGCA CCGTTCTCAC TGCATTTCATG GCGGATATCT CAGATATCCG GGTGNAACGA AGCAATCTTG CCGCGATGTT AATTATGTAT

```

15/35

FIGURE 5B

801 ACCTTATGTA TCATACACAT AGATTTTAGG TGACACTATA GAATACATC CACTTTGCCT TTCTCTCCAC AGGTGCCAC TCCCAGGTCC AACTGCACCT
TGGATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTATTGTAG GTGAACCGA AAGAGAGGTG TCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCCGGGATC CTCTAGAGAT CCCTCGACCT CGAGTCGACT TTTTTTTTTT TTTTGTAGG CCAAGGGTA CTTCTTTTTT
GCCAAGATAG CTAACTTAAG GGGCCCCCTAG GAGATCTCTA GGGAGCTGGA GCTCAGCTGA AAAAAAANA GGTTCCTCCAT GAAGAAAAAG

1001 TTTATTAATT ACTCAGAAGT CTAGGCCACA GCAATCTACT GTTCTCTCTCT CATTTTCTTA AACTATTTTG ATACCTAATT CTCAGACTTT ATGGGCTATT
AAATAATTAA TGAGTCTTCA GATCCCGTGT CGTTAGNTGA CAAGAGGAGA GTAAAGGAT TTGATAAACC TATGGATAAA GAGTCTGAAA TACCCGATTA

1101 AGACATTTCT CACATTTCCA TAGATAATTA CTCATCCGTT TTGCAACCTG ATTCTCAATA TTAAGAGATT AAACTAATG TATATGACTC TCAGTTGACA
TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGGCA AACGTTGGAC TAAGAGTTAT AATTCTCTNA TTTTGATTAC ATATACTGAG AGTCMACTGT

1201 CATACTGAAG TACAGAAAAA TTCCATCAAT TCCTTCTGCA AATGAAAAA GACTTCGTTT TCTCAACAGC TGCACTAATT TTTTATGCAT AGAAAAAAT
GTATGACTTC ATGCTTTTTT AAGTAGTAA AGGAGACCT TTTACTTTTT CTGAAGCAAA AGAGTTGTG ACGTAGTAAA AAAATACGTA TCTTTTTTA

1301 GTGCAATTAC TCCAAGTACA ATCAAGTCAT TTAACATGSC TTTACCATCA TTGTAGTTAC AGGATATTTT AAAAGAGAAA AAAAATCTC AAAGCACAGG
CACGTTAATG AGGTTCAATG TAGTTCAGTA AATTGTACCG AATGGTAGT AACATCAATG TCCTATAAAA TTTTCTCTTT TTTTGTAGG TTTCTGTCTC

1401 TCCTGCTGTG CAGCAAGCA ATCAAAATCC TTCAATAATA CAGCCTGATG GGATTCAGCA ATCTGAGGAA TAATGAATTA CCACCTTAAT CAGTAAACAG
AGGACGACAC GTCGTTTCTGT TAGTTTAAGG AAGTATTATT GTCGGACTAC CCTAAGTCGT TAGACTCCTT ATTACTTATT GGTGAGATTA GTCAATTGTC

1501 GAATATGCTA CAACAGTCAC TGAGTAAAAA TTGCACTATC ATCTGTTGAT TCTCTTGATC GACATTTCAA ACAATAAATG GAAATGTAAG TATCTCTTAA
CTTTTACCAT GTTGTCACTG ACTCATTTTT AACCTGATAG TAGCAACTA AGAGAACTAG CTGTAAAGTT TGTATTATAC CTTTACATTC ATAGACAAAT

16/35

FIGURE 5C

1601 AAAGAAAAT AACTGGGTTT AGTGTGCTTA ATTTTACCAG GCAGTGAGGA AATTATATAT CACCTTGACT GTCTGCAGT GTTGCCCACT CAATAAATG
 TTTCTTTTAA TTGNAACCAA TCACACGAAT TAAATGGTC CGTCACTCCT TTAATATATA GTGGAAGTGA CAGGACGTCA CAACGGGTCA GTTATTTTAC

1701 CACAAATAAT CTTTTTCATA ATACATGGCC AACTTTATCC TATCAGTTGA ATATGTCAGG ATAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG
 GTGTTTATTA GAAAAAGTAT TATGTACCGG TTGAAATAGG ATAGTGAAGT TATACAGTCC TATTTGACTA ACACGTCNAC CAACATTTGT AACATNAAC

1801 GAATGGAATTA TTTGAATTTG TTTTGCTACT TTTTGCTACT ATATTCTTCT CCAAGTGTTC TCATTGAAGG TTATTGTCAT CTGAATATGA AGAGTCTGTT
 CTTACCTAAT AACTTAAAC AAAAGGNTGA AATAATNAAC TATAAGAAGA GGTCACTAAGT AGAATCTTTC AATAACGTA GACTTATACT TCTCAGACAA
 506 O R I F N N A D S Y S S D T

1901 TCAAAATAGT CTTCAAGTTT CCAACGCAGT GTCTCAATG TAGTTCGTTT CTTAGGCTCT GCATCCAGC ACTCCAACAT GATGTTGTAA AATTGCTGTG
 AGTTTATCA GAAGTCAAA GGTTCGTCA CAGAGTTTAC ATCCAGCAAG GAAATCCGAGA CGTAGGTGTA CTACAACATT TTAACGACAC
 492 E F Y D E L K W R L T E F T P R E A N W C E L M I N Y F Q Q

2001 GACAGTTGGA TGGTTGCGGA AGTCTATAGT TTTGAGCCAA CATCTGANTT ACCTGGGCAC CTGTCAATACC ACTGTAAGGC ATTTTGCCAT AAGTAATGAT
 CTGTCAACCT ACCAAGCCT TCAGATATCA AACTCGGT GTACACCTAA TGGACCCGTG GACAGTATGG TGACATTCCG TAAACGGTA TTCATTTACTA
 459 P C N S P Q P L R Y N Q A L M Q I V Q A G T M G S Y P M K G Y T I I

2101 TTCATNAAGA AGGATTCCAA ATGACCATAC ATCGGACTTA ATGCTGANTT TATTACTACG AATGGCTTCG GGCGCAGTCC ACTTCACCGG CAGCTTTATT
 AAGTATTTCT TCCTAAGGTT TACTGGTATG TAGCCTGAAT TACGACTTAA ATATGATGC TTACCGAAGC CCGCGTCAGG TGAAGTGCC GTGGAATAA
 425 E Y L L I G F S W V D S K I S F K N S R I A E P A T W K V P L K I

2201 TCGTGTCTAG ATTATAGAT GTCTTCATTA TCTACCTTAA AACTCTGGC AAGTCCAAA TCTGCTACTT TGATGATATT ATGTTCAACA ACGAGGACNT
 AGCACAGATC TAAGTATCTA CAGAGTAAT AGATGGAAT TTTGAGACCG TTCAGGTTTT AGACGATGAA ACATCTATAA TACAAGTGT TGCTCTCTGA
 392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGGCAGC CAGATCTCTG TGAATGTAGT TCCGAGACTC CAGATAGGCC ATTCCAGAGG CAACCTGTGC GGCATGTCT ACCTGTTGAG TCAGATGGAT
 AAGACCGTCG GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCG GTTGACACG CCGGTACAGA TGGACAACTC AGTCTACCTA
 359 N R A A L D R H I Y N R S E L Y A M G S A V Q A A M D V Q Q T L H I

17/35

FIGURE 5D

2401 TTTTGGATCCA GTGTCAATTTT GGAGATATTC TTGCAGACTT CCATGCTCA TCAACTCTGT AATAATATAA ATTGGATCTT CTAAGTGCA AACAGATTA
 AANACTAGGT CACAGTAATA CCTCTATAAG AACGTCTGAA GGTACAGAGT AGTTGAGACA TTATTATATT TAACTTAGAA GATTTCACGT TTGTCTGATT
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y

2501 AGCTGGATAA GCTTGGGATG TCTTAGGTTCT TCTATTATCT GTGCTCCCT CAGGAAGTCA TTTGGATCCA TTGAACCTGG TTTTAATGTT TTCACCTGTA
 TCGACCTATT CGAACCTTAC AGAATCCAAG AAGTAATAGA CACGGAGGGA GTCTTTCAGT AAACCTAGGT AACTTGGACC AAAATTACAA AAGTGACGAT
 292 L Q I L K P H R L N K M I Q A E R L F D N P D M S G P K L T K V A

2601 CTGGAGTGGT ATGTTCAC AGACCTTCCC ATACTTCGCC AACTGACCA GATCCCAATC GCTTCAGAAG CTGTATGGAG TTGGGGTCTA TCTCCCATTTG
 GACCTCACCA TAACAAGGTG TCTGGAAAGG TATGAAGCGG TTGACTGGT CTAGGGTTAG CGAGTCTTC GACATACCTC AACGCCAGAT AGAGGGTAAC
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q

2701 GTCCACGGTT TTATACGACA ATCAAAATGG AGCTGGGACC TGGATCTTGA AGCATGGTTT CCCAGCTTG ACACACAGGC CGTCACTTGT CTTGGTGTAG
 CAGGTGCCAA AATATGCTGT TTAGTTTACC TCACCCCTGG ACCTAGAAAT TCGTACCAAA GGGTCCGAAC TGTGTGTCG GCAGTGAACA GAACCATATC
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y

2801 TGGCTACAA ATTCTGTCAG TGTGNAAG ATTCTTCTTC GCGTGAGAA AATCCCTCT TCATCCAGTC TTTTAATCT GTAGTGTCTT ACAACTGCTC
 ACCGAGTGT TAAGCAAGTC ACAACTTTC TAAGNAGAG CGCACTCTTT TTTAGGGGA AGTAGGTGAG AAAATTAGA CATCACAAA TGTTCACGAG
 192 H S V F E N L T S F I R R R T L F F G G E D L R K I R Y H K V V A

2901 CATCTAAAC TGAAGAGAG AATCTCTCTT TTTGGCTTTC ACTTCTCTG ATTAGAAAG AACCGTCTT GTTTCCTGAA TATAATAGTT GTTTCCTGCG
 GTAGATTTTG ACTTCTCTC TTAGAGGAA AACCCGAAAG TGNAGAGAC TAATCTTTC TTGGCCAGAA CAAAGAGCTT ATATTATCAA CAAAGAGACG
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A

3001 ATCTGATCTT CCGATTGCTC CAAAGAACCA CGGCTCTGCC TGTAGGCTTC TGTCTCAGC CACGTAGTTA GAAGGAATAT AGCCTTGTAG TTGCTGACTG
 TAGACTAGAA GGCTAACGAG GTTCTTGTG GCCGAGACGG ACATCCGAG ACAGGAGTGG GTGCATCAAT CTTCCTTATA TCGGAACATC AACGACTGAC
 125 D S R G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S

3101 GAGCCATCTC GTCTTTTCTC CAAGTGTCTG GCAMCCACC AGCCCTCMTG CAAAGTGTC AGAATTGAA GTTTGTACC TGCTCGGAG CTCGAAGTCTC
 CTCGGTAGAG CAGAAAGAG GTTCACAGAC CGTTTGGTGG TCGGAGTAC GTTTCACAGG TCTTGAACCT CAAACAGTGG ACGAGCCTTC GAGTTCAGGA
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

FIGURE 5E

3201 CAGCAGTCCG AGCCTGGTAA TCACAAAG-CCACAAAGTA GTGGCCATGC CTCGTGACT GGGGAGAGCA AAGGGCCCTT GGATTTTCAA TCACGGTTGA
 GTCGTCAGGC TCGGACCAT AGTTTGTTC GGTGTTTCAT CACCGGTACG GAGACACTGA CCCCTCTCGT TTCCCGGGGA CCTAAAAGTT AGTCCCACT
 59 E A T R A Q Y D F L A V F Y H G H R Q S Q P S C L A G P N E I V T S

3301 CTTGTCTGCC TCCGTGGACA AACAGGGGAG ATAGGTTTCT AGTACTCCC AGAGCCTCTG ACAGATGTTG CTCATTGTGC CTTGGTGGGG AGAAGAGGAG
 GAACAGACGG AGGCACCTGT TTGTCCCTC TATCCCAAGA TCCATGAGGG TCTCGGAGAC TGTCTACAAC GAGTAACACG GAACCACCCC TCTTCTCTCTC
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

3401 CAGGGCTTCT CCTCTCCCC TTAGTCTCTG CGATCCACCT TATCTTCTT CACCAGGCAA CTTTGAAGTC AGCACCACCT CACCATACTT CGGAGAGTAT
 GTCCCGNAGA GGGAGAGGGG AATCAGAGAC GCTAGGTGGA ATAGAAGGNA GTGTCCGTT GAACTTCAG TCGTGGTTGA GTGGTATGAA GCCTCTCAT

3501 GCAAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTTGCA GCAAGTCTTA CTTACCGGCT TGCTTTCTGT GGCTGGAGGT GCTACCCCGA
 CGTTTCAGGG CAAAGTCTAG TCAGGTCTAG GACCCAACGT CGTTCAGAT GGACCTCTCT GAATGGCCGA ACGAAGACA CCGACCTCCA CGATGGGGCT

3601 GGCNAACCTG AGCAGGAGCT GGGCAGCTGC TCACTAGGAA GGTGTCTTTT CTCTTATCT GCTTAAAGAT CCCACAACAA AATAAANAATA AATTTAAAAG
 CCGTTTIGAC TCGTCTCTGA CCGGTCTCGA CCGGTCTCGA AGTGATCTT CCACAGNAAA GNAGATAGA CGAATCTTA GGGTGTGTGT TTTATTTTAT TTTAATTTTC

3701 GGCTTTATTT AGACAATAT CTGAGNACAG AATGGTGCCA TCTTGCTTT TGTCCCAATA AAGATTAGC AAGAGGAAGC TACTAACCCC TGGTAAAACC
 CCGAATATA TCTGTTTATA GACTCTTGTG TTACCACGGT AGNACGNA ACAGGGTTAT TTTCAATCG TTCTCTCTCG ATGATTGGGG ACCATTTTGG

3801 TCCACGTCTT GCTTTCGCCA GGTTCGACTC GAGGGATCTT CCATACCTAC CAGTTCTGCG CCGGCGCGGA CTCTAGAGTC GACCTGCAGA
 AGGTGCAGAA CGAAGCGGT CCCAGCTGAG CTCCCTAGAA GGTATGATG GTCAAGACGC GGACGTCCAG CGCCGGCGGT GAGATCTCAG CTGGACGTCT

3901 AGCTTGGCCG CCATGGCCCA ACTGTGTTTAT TGCAGCTTAT AATGGTTACA AATAAGCAA TAGCATACA AATTTCACAA ATAAAGCATT TTTTTCACGT
 TCGNACCGGC GGTACCGGGT TGNACNANTA ACGTCGNATA TTACCATGT TTAATTCGTT TTAAGTGTT TATTTCGTAA AAAAAGTGAC

19/35

FIGURE 5F

4001 CATTCTAGTT GTGGTTTGT CAACTCATC AATGTATCTT ATCATGTCTG GATCGGAAT TAATTGGCG CAGCACCATG GCCTGAATA ACCTCTGAAA
 GTAAGATCAA CACCAACAG GTTTGAGTAG TTACATAGAA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCGTGGTAC CGGACTTTAT TGGAGACTTT

4101 GAGGAACCTG GTTAGGTACC TTCTGAGCG GAAAGMACCA GCTGTGGAAT GTGTGTCACT TAGGGGTGG AAAGTCCCCA GGCTCCCCAG CAGGCAGAAG
 CTCTTGNAC CAATCCATGG AAGACTCCGC CTTTCTTGGT CGACACCTTA CACACAGTCA ATCCACACC TTTCAGGGGT CCGAGGGGTC GTCCGTCTTC

4201 TATGCAAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA ACTATGCMAA GCATGCATCT CAATTAGTCA
 ATACGTTTCG TACGTAGAGT TAATCAGTCTG TTGTCTCACA CCTTTCAGGG GTCCGAGGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGCCC TAACTCCGC CAGTTCCGCC CATCTCCG CCGATGGCTG ACTAATTTTT TTTATTTATG
 CGTTGGTATC AGGGCGGGA TTGAGCGGG TAGGGCGGG ATTGAGCGG GTCAAGCGG GTAAGAGCG GGTACCCGAC TGATTAATAA AAATAAATAC

4401 CAGAGGCCGA GCGCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTG GAGGCTTAGG CTTTTCMAA AAGCTGTAA CAGCTTGCCA
 GTCTCCGGCT CCGCGGAGC CGGAGACTCG ATTAGTCTT CATCACTCTT CCGAATAAAC CTCGGATCC GAAACGTTT TTCGACAATT GTCGAACCGT

4501 CTGGCCCTCG TTTTACAACG TCGTGACTGG GAAACCCCTG GCGTTACCCA ACTTAATCGC CTTCGACGAC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG
 GACCGGCAGC AAATGTTGC AGCACTGACC CTTTGGGAC CGCAATGGT TGAATTAGCG GAACGTCTG TAGGGGAAA GCGGTGACC GCATTTATCGC

4601 AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCCGAG CTTGAATGGC GAATGGCGCC TGATGGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTT
 TTCTCCGGC GTGGCTAGCG GGAAGGTTG TCAACCGTC GGACTTACCG CTTACCGCG ACTACGCCAT AAAAGAGAA TGCCTAGACA CGCATATAAG

4701 ACACCGCATA CGTCAAAAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGCG GTGTGTGTG TTACGGCGAG CGTGACCGCT ACCTTGCCA
 TGTGCGTAT GCAGTTTCGT TGGTATCATG CGCGGACAT CGCGCGTAA TTGCGCGCG CCACACCACC AATGCGCGT GCACTGGCGA TGTGAACCGT

20/35

FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTCGCTTTCT TCCCTTCCTT TCTCGGCACG TTCGCGGGCT TTCCCCCGTCA AGCTCTAAAT CGGGGGCTCC CTTTAGGGTT
CGCGGGATCG CGGGCGAGGA AGCGGAAAGA AGGGMAGGA AGAGCGGTGC AAGCGGCCGA AAGGGGCACT TCGAGATTTA GCGCCCGAGG GAAATCCCA

4901 CCGATTTAGT GCTTTACGGC ACCTCGACCC CAAAAACTT GATTGGGTG ATGTTACG TAGTGGCCA TCGCCCTGAT AGACGGTTT TCGCCCTTTG
GGCTAAATCA CGAAATGCCG TCGAGCTGGG GTTTTGTGA CTAACCCAC TACCAAGTGC ATCACCCTT ATCACCCTT AGCGGACTA TCTGCCAANA AGCGGNAAC

5001 ACGTTGGAGT CCAGTTCTT TAATAGTGA CTCTGTCTC AACTGGAAC AACACTCAAC CCTATCTGG GCTATCTTT TGATTTATAA GGGATTTTGC
TGCACCTCA GGTGCAAGAA ATTATCACCT GAGAACNAGG TTTGACCTTG TTGTGAGTTG GGATAGAGCC CGATAAGANA ACTAATATT CCTAANAACG

5101 CGATTTCGGC CTATTGGTTA AAAATGAGC TGATTTAACA AAATTTAAC GCGAATTTTA ACAATATTT AACGTTTACA ATTTTATGTT GCATCTCAG
GCTAAAGCCG GATAACCAT TTTTACTCG ACTAATTTG TTTTAATTT CGCTTAAAT TGTTTATAA TTGCAATGT TAAATATCCA CGTGAGAGTC

5201 TACAATCTGC TCTGATCCG CATAGTTAAG CCAGCCCCGA CACCCGCTGA CCGGCTTGC CGGCTTGTG TGTCCCGGC ATCGCTTAC
ATGTTAGAC AGACTACGG GTATCAATTC GGTGGGGCT GTGGCGGCT GCGCGGACT GCGGGAACAG ACGAGGGCCG TAGGCGAATG

5301 AGACAAGCTG TGACCGTCTC CGGAGCTGC ATGTGTCAGA GGTTTTCACC GTCATCACCG AAACGCGGA GACGAAAGG CCTGTGATA CGCTATTTT
TCTGTTGAC ACTGCGAGG GCGCTCGAG GCGCTCGAG TACACAGTCT CCAAAAGTGG CAGTAGTGG TTTGCGGCT CTGCTTTCC GGAGCACTAT GCGGATANA

5401 TATAGGTTAA TGTATGATA ATANTGTTT CTAGACGTC AGTGGCACT TTTGCGGGA ATGTGCGCG AACCCCTATT TGTTTATTT TCTAATATACA
ATATCCAAT ACAGTACTAT TATTACCANA GAATCTGCAG TCCACCGTA AAAGCCCCCT TACACGCGC TTGGGATTA ACAAATANA AGATTATGT

5501 TTCAATATG TATCCGCTCA TGAGACAATA ACCCTGATA ATGCTTCAAT AATATTGAA AAGGAAGAT ATGATATTC AACATTCCG TGTGCCCCCT
AAGTTTATAC ATAGCGAGT ACTCTGTTAT TGGGACTATT TACGAAGTTA TTATAACTTT TTCTTCTCA TACTCATAG TTGTAAAGG ACAGCGGGA

21/35

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCTTTCCT GTTTTGTCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT
 TNAGGGMAAA ACGCCGTAA AACGGGTA AAGCGGTAAGG CAAVAMCGAG TGGGTCTTTG CGACCACTTT CATTTTCTAC GACTTCTAGT CAACCCACGT GCTCACCCAA

5701 ACATCGAACT GGATCTCMAAC AGCGGTAAGA TCCTTGAGAG TTTTCGCCCC GAAGNACGTT TTCCAAATGAT GAGCAGCTTTT AAAGTTCTGC TATGTGGCGC
 TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAATCTTC AAMGCGGGG CTTCTTGCAA AAGTTACTA CTCGTGAAA TTCAAGACG ATACACCGCG

5801 GGTATATCC CGTATTGACG CCGGGCAAGA GCAACTCGT GCGCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAANAAGCAT
 CCATAATAGG GCATAACTGC GGCCCGTTCT CGTTGAGCCA GCGCGGTATG TGATAAGAGT CTTACTGAAC CAATCATGA GTGTCAGTG TCTTTTCGTA

5901 CTTACGGATG GCATGACAGT AAGAGAATTA TGCAGTGTG CCATAACCAT GAGTGATAAC ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA
 GAATGCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG TGACGCCGT TGAATGAAGA CTGTTGCTAG CCTCTGGCT

6001 AGGAGCTAAC CGCTTTTGTG CACAACATGG GCGATCATGT AACTCGCTT GATCGTTGG AACCGAGCT GAATGAAGCC ATACCAACG ACAGCGGTGA
 TCCTCGATTG GCGAANAAC GTGTGTATAC CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGCGCTCGA CTTACTTGG TATGTTTGC TGCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAC GTTGGCGCAA CTATTAACTG GCGAACTACT TACTCTAGCT TCCCGGCAAC AATTAAATAGA CTGGATGGAG
 GTGGTGCTAC GGACATCGTT ACCGTGTGTG CAACGGTTT GATNAATTGAC CGCTTGATGA ATGAGNTCGA AGGCCCGTTG TTAATTATCT GACCTACCTC

6201 GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATGCT GATAAATCTG GAGCCGTTGA GCGTGGTCT CGCGGTATCA
 CGCCTATTTC AACGTCCTGG TGAAGACGG AGCCGGNAG GCCGACCGAC CAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAGCACT GGGGCCAGAT GGTAAGCCCT CCGTATCGT AGTTATCTAC ACGACGGGA GTACAGGCAAC TATGGATGAA CGAAATAGAC AGATCGCTGA
 AACGTCGTGA CCCCCTCTA CCAATCGGGA GGCCATAGCA TCAATAGATG TGCTGCCCT CAGTCCGTTG ATACCTACTT GCTTTATCTG TCTAGCGACT

22/35

FIGURE 51

6401 GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAACTTTACT CATATATACT TTAGATTGAT TTAAAACTTC ATTTTAAAT TAAAGGATC
CTATCCACGG AGTGAATAAT TCGTAACCAT TGACAGTCTG GTTCAATGA GTATATATGA AATCTAACTA AATTTGAG TAAAAATTA ATTTTCTTAG

6501 TAGGTGAAGA TCCTTTTGA TAATCTCATG ACCAATCC CTTAAGTGA GTTTTCGTC CACTAGCGT CAGACCCCGT AGAAAAAGATC AAAGGATCTT
ATCCACTTCT AGGAAAAACT ATTAGAGTAC TGGTTTATAG GAATTGCACT CAAAAGCAAG GTGACTCGCA GTCGCGGCA TCTTTTCTAG TTTCTTAGA

6601 CTTGAGATCC TTTTTCCTG CGCGTAATCT GCTGCTTGA AACAAAAA CCACCGCTAC CAGCGTGT TGTGTCGG GATCAAGAGC TACCAACTCT
GAACTCTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG GTGCGCACCA AACAAACGGC CTAGTTCTCG ATGTTTGAGA

6701 TTTTCCGAG GTAACGTGGT TCAGCAGAGC GCAGATACCA AATACTGTC TTCTAGTGA GCCGTAGTGA GGCACCACT TCAAGAACTC TGTAGCACCG
AAAAGCTTC CATTGACCGA AGTCGTCTCG CGTCTATGGT TTATGACAAG AAGATCAGAT CGGCATCAAT CCGGTGTGA AGTTCTTGAG ACATCGTGGC

6801 CCTACATACC TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGCTTTACCG GGTGGAATC AAGACGATAG TTACCGGATA
GGATGTATGG AGCGAGACGA TTAGGACAAT GGTCAACCGAC GACGGTCACC GCTATTTCAGC ACAGATGGC CCACCTGAG TTCTGCTATC AATGGCTAT

6901 AGCGCGACCG GTGCGGCTGA ACGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA
TCCCGGTGCG CAGCCCGACT TGCCCCCAA GCACGTGTGT CCGGTGGAAC CTCGCTTGT GATGTGTATG GATGTGCGAC TCGATACTCT

7001 AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGTGCG AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC
TTCCGGGTGC GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTGCG CGTCCAGCC TTGTCCTCTC GGTGCTCCC TCGAAGGTCC CCTTTGCGG

7101 TGGTATCTTT ATAGTCTGT CGGTTTTCG CACCTCTGAC TTGAGCGTGG ATTTTGTGA TGCTGCTCAG GGGGGCGGAG CCTATGGA AAACGCCAGCA
ACCATAGAAA TATCAGGACA GCCCAAGCG GTGGAGACTG AACTCGAGC TAAAAACACT ACAGCAGTC CCCCCGCTC GGATACCTTT TTGCGGTCTG

FIGURE 5J

7201 ACGCGGCCCTT TTTACGGTTC CTGGCCCTTT GCTGGCCTTT TGCTCACATG TTCTTTCCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT
TGCGCCGGAA AAATGCCAAG GACCGGAJA GACCGGAA ACGAGTGAC AAGAAGGAC GCAATAGGG GCAATAGGAC CTATTGGCAT AATGGCGGNA

7301 TGAGTGAGCT GATACCGCTC GCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGA GAGCGCCAA TACGCAAAACC GCCTCTCCCC
ACTCACTCGA CTATGGCGAG CCGCGTCGSC TTGCTGGCTC GCGTCGCTCA GTCACTCGCT CCTTCGCCCT CTGCGCGGT ATGCGTTTGG CGGAGAGGGG

7401 GCGCGTTGGC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCAAT
CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGCTGAC CTTTCGCCG TCACTCGGT TCGTTAAT ACACCTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCCTTACAC TTTATGCTTC CGGCTCGTAT GTTGCTGGA ATTGAGCG GATAACNAAT TCACACAGGA AACAGCTATG ACATGATTAC
TCCGTGGGT CCGAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA AGTGTGCTT TTGCGATAC TGTACTAATG

7601 GAATTAA
CTTAAT

24/35

1 GCGGCCGCAG AGAAGCAGA GGATGGGGCT TAGCAGCTGG CAGAGCCAGG ACCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGNAGTC CTGAACCTTT
CGCCGGGGTC TCTTTCGTCT CTTACCCCGA ATCGTCGACC GTCTGGTCC TGGCCCCCTCC ATCGTCTTTC TCGTGTTCAT GTTCTTCAG GACTTTGAAA

101 GGTTTTGGTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAGACCC TGAAGATCAC CGACTTTGGC CTGSCCCGAG ACTGGCACAA AACCCACAAA
CCAAACGAC GACGTGGGT AACTCTACT GCTGTACCTC GTGTCTGGG GTGGAACCG GACCGGGCTC TCACCGTGTI TTGGTGTGT

201 ATGAGTGCCG CXGGCACCTA CXCTGGATG GCTCCTGAGG TTATCAAGGC CTCACCTTC TCTAAGGGCA GTGACGTCTG GACTTTTGGG GTGCTGCTGT
TACTCACGGC GPCCGTGGAT GPGGACCTAC CGAGGACTCC AATAGTCCG GAGGTGGAAG AGATTCCCGT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GGGAACTGCT GACCGGGGAG XTGCCATACC GTGGCATTGA CTGCCCTGCT GTGGCCTATG GCGTAGCTGT TAACAAGCTC ACAGTCCCAT CCATCCACCT
CCCTTGACGA CTGGCCCTC PACGGTATGG CACCGTAACT GACGGAAACGA CACCGATAC CGCATCGACA ATTGTTCGAG TGTACGGTA GGTAGGTGGA

401 GGCC
CCGG

FIGURE 6

25/35

FIGURE 7A

1 ATGAGAGGCT TGGGCGGCA GGGGGGAG CTGGGCTGG TGGTGTGTTT TTATGCAATG ATATTGCGA CTATTACAAA TCAAGATGCTG GCTGATGATCA
TACTGTGCA ACCGGGCTT GGGGGGCTG GAGGGGAG AGCAATAAAA AGAGGCTTAC TATAAAGCTT GATAATGTTT AGTTCTAGAG GGACACTAGT

101 AGTGTGTTTT AATCAATCAT AAGAACAAATG ATTATGAGT GGGAACTGA TTATCATATC CCATGCTATC AGATCCGCG GAGACACTG GGTGTGCTTT
TCACACAAAA TTACTTACTA TTCTTTCTAG TAAGTACTGA GGGTTCAGT ACTATATAG GGTACCATAG TCTTAGGCGC CTTCGCGAG CCACACCGCA

201 GAGACCCGAG AGCTCAGGA CAGTCTACCA AGCTGCTCT GTGGAAGTGG ATCTATCTGC TTCCATCACA CTGCAAGTGC TGGTCGATGC CCCAGGGGAG
CTCTGGGTC TCGAGTCCCT GTCACATGCT TCGACGGGCA CAGCTTCAGG TACATAGAG AGGTAGTGT GAGCTTCAG ACCAGCTAGG GGGTCCCTTG

301 ATTGCTGTC TCTGGTCTT TAAGCACAG TCCCTGAATT GCGAGGACCA TTTTATTTA CAAACAGAG CAGTGTTC CATGCTCAT TTCAAAATGA
TAAGGACAG AGACCGAG AAAGCTGCG AGGACTTAA GGTGCTGCTT AAAGTAAAT GTTGTCTC CTCACAGAG GTACGCTAA ACCTTTACT

401 CAGAAACCCA AGCTGAGAA TACCTACTTT TTATTACAG TGAAGTATC AATTAACAA TATTGTTTAC ATGATATATA AGAAATACC TGGTTTACAC
GTCTTGGT TCGACCTTT ATGATGAA AATAAGTGC ATTGATGAG TTAATGTTTT ATACAAATG TCACTCATAT TCTTTATGAG ACIAAATGTG

501 ATTAAGAAGA CTTTACTTTA CAAAATGGA AAACGAGGAG GCGCTGGTCT GCATATCTCA GAGGCTTCCA GAGCGGATCC TGGAAATGGT GCTTTGGCAT
TAATCTTCT GAAATGAAT CTTTTACT TTTGGTCTG GGGACAGA GGTATAGACT CTCGAGGT CTCGCTAGG ACCTTACACA CGAAACGCTA

601 TCACAGGGG AGAGCTGAA AGAAGAAAT CCAGCTGTTG TTAAAGGA GGAAGAGT CTCATGAAT TATTTGGAC GGACATAAG TGTGTGTGCA
ACTGTCCCG TTGAGCAT TCTTCTTTA GGTGACAG AATTTTCT CTTTTTCAG GAAGTACTTA ATAAAGCTG CCGTATTTCC AGGACAGGCT

701 GAAATGAAT GGGCAGGGA TGCACAGG TGTTCATTAAT AGATCTAAT CAAAGCTG AGAGCAGAT GGCACATTA TTTCTTAAG TAGGGAACC
CTTAATTTA GGTGCTT ACTGTGAG ACAATCTTA TCTAGATTTA GTTGAGAG TGTGCTTAA GGTGTTAAT AAGGATTTG ATGCTGCTTG

26/35

FIGURE 7B

801 CTTATGATGA AGGTGCAAG CTGTTGATCT GAACCATGGA TTGCGGCTCA CCTGCGAATT AGAAGACAAA GCACTCGAG AGGGCACTA CTTTCAGATG
GAATACCTAT TCCAGCTTC GACAAGTACA CTGCTACTCT AAGCCGAGT GAACCCCTAA TCTTTTCTTT CGTGAGCTCC TCCCGTTGAT GAAACTCTAC

901 ACTACCTATT CAACAACAG AACTATGATA CGGATTCTCT TTGCTTTTCT ATCATCACTG GCAAGAAACG ACACCGGATA CTACACTTGT TCCTCTTCAA
TCATGATATA GTTGTCTGTC TTGATACTAT GCTTAAGTA AAGCAAAACA TAGTAGTCAC CTTTCTTTTC TGTGGCCTAT GATGTGAACA AGGAGAAGTT

1001 AGCATCCGAG TCATCAAGCT TTGCTTACCA TCGTAGAAAA GCGATTATA AATGCTACCA ATTCAAGTCA AGATTATGAA ATTGACCAAT ATGAAGAGTT
TCGTAGGCTC AGTTAGTCCA AACCAATGCT AGCATCTTTT CCTAAATAT TTACGATGCT TAAGTTCACT TCTAATACCT TAACTGCTTA TACTTCTCAA

1101 TTGTTTTTCT GTGAGGTTTA AAGCTACCC ACAATACGA TGTACCTGGA CCTTCTCTCG AAATCATTT CTTGTGAGC AAGGGTCT TGTATACGGA
AACAAAAACA CAGTCCAAAT TTGCGATGCG TGTTAATCT ACATGACCT GAAACAGAG TTTTAGTAAA GGAACACTCG TTTTCCCAAG ACTATTGCT

1201 TACAGCATAT CCAAGTTTTG CAATCATAG CACCAAGCAG GAGATATAT ATTCCATGCA GAATATGATG ATGCCCAAT TACCAAAATG TTCAGGCTGT
ATGTCGTATA GTTCAAAAC GTTAGTATTC GTGCTGCTC CTCCTATATA TAAGGTAGCT CTTTACTAC TACGGTTAA ATCGTTTTAC AAGTCCGACA

1301 ATATAGAAG GAACCTCMA GTGCTCGCAG AAGCTCGGC AATCAAGCG TCGTGTCTCT CGGATGGATA CCCATTACCA TCTTGAGCT GAAAGAGTG
TATATTCTTC CTTTGAATT CAGGAGCTTC TTGGAAGCG TTCACTCGC AGGACAAAGA GCTACTAT GGTAAATGCT AGAACCTGGA CTTTCTTCAC

1401 TTGACACAG TCTCCCACT GCACAGAGA GATCACAGAA GACTCTGGA ATACAAAGAG TAACAGAAAA GTGTTTTAG AGTGGTTC GAGCACTACT
AAGTCTGTC AGAGGTTGA CGTGCTCTCT CTACTGCTCT CTTAGAGCT TATCTTTCCG ATTGCTCTT CACAAAGCTG TCACCCACAG CTCGTCTATG

1501 CTAAACATGA GTGAAGCCAT AAAGGGTTC CTGCTCAACT GCTGTGCAAT CAATTCCTCT GGCACATCTT GTGAGAGCAT CTTTTAAG TCCTCAGGCC
GATTTCTACT CACTTCGTA TTTTCCCAAG GATCAATTA CAGACAGTAT GTTAAGGGA CCGTGTAGA CACTCTGTA GGAATTTG AGAGTCCG

27/35

FIGURE 7C

1601 CTTTCCCTTT CATCCAGAG ACATCTCAT TCTATGCAAC AATTGGTGT TGTCTCTCT TCAATTGCGT TTAAACCCCTG CTAATTTGTC ACAAGTACAA
 GCAAGGCAAA GTAGGTCTG TTGTAGATA ACATAGCTTG TTAAGCAACA ACAGAGGACA AGTAACAGCA AATTTGGGAC GATTAACAG TGTTCATGTT

1701 AAAGCAATTT AGGTATGAAA GCGAGCTACA GATGCTACAG GTACCGGAT CCTCAGATAA TGAGTACTTC-TACGTTGATT TCAGAGATAA-TGAATATGAT
 TTTGCTTAAA TCCATACTTT GGTGCTATCT CTACCATGTC CACTGCTCTA-GGAGTCTATT-AGTCATGAG ATGCAACTAA AGTCTCTTAT ACTTATACTA

1801 CTCAAATGGG-AGTTTCCAAAG AGAAATTTTAAGATTTTGGGA AGTACTAGG ATCAGGTGCT TTGGAAAAG TCATGAAAGG AACAGCTTAT GGAATTAGCA
 GACTTTACCC TCAAGGTTTC TCTTTTAAAT CTCAAACCCCT TCCATGATCC TAGTCCACCA AACCTTTTC ACTACTTCCG TTGTGGAATA CCTTAATCGT

1901 AAGCAGAGCT-CTCAATCCAG GTTACCCTCA; AANTGCTGAA; AGAAAGCA GACAGCTCTG AAGAGAGGG-AGTCATGTCA GAACCTAAGA-TGATGACCC-
 TTTGCTCTCA GACTTAGGTC CAATGGCAGT TTTACGACTT TCTTTTTCGT CTGTGAGAC TTTCTCTCCG TGAGTACAGT CTTGAGTTCT ACTACTGGGT

2001-GCTGGGAGG-CACAGATAA TTGTGAAGCT-GCTGGGGGG TGCACACTGT CAGGACCAAT TTACTGATT TTTCATATCT-GTGTCTATGG-TGATCTTCTG
 CGACCCCTTC GTGCTCTTAT AACACTTGGG CGACCCCGCG AGCTGTGACA GTCTGTGTTA ATGCAACTAA AACTTTATCA CAACGATACC ACTAGAAAGAG

2101-AACTATCTAA-GAAGTAAAG AGAAATTTT-CACAGGACTT GACACAGAT TTTCAGGAA CACAATTTCA GTTTTACC-CACTTTCCAA TCACATCCAA
 TTGATAGATT CTTCATTTTC TCTTTTAAA GTCTCTGAA CTGTCTCTA AAGTTCTCT GTGTTAAGT CAAAATGGG GTGAAGGTT AGTGTAGGTT

2201-ATTCAGGAT-GCCTGGTTCA AGAGAAGTTC-AGATACACC GCACTGGAT CAATCTCAG GGCCTCATGG GAATTCATTT CACTCTGAAG ATGAATTTGA
 TTAGTCTGTA CCGACCAAGT TCTCTTCAAG TCTATGTGG CCTGAGGCTA GTTAGAGTC CCGAGTACC CTTAGTAAA GTGAGACTTC TACTTTAACT

2301 ATATCAAAAC CAAAAGGC TCGAAGAGA GGAAGGACTTG AATGTGTTA CATTGAGA TCTTCTTTGG TTGCTATTC AGTTGCCAA AGCAATGCAA
 TATACTTTTG GTTTTTCGG ACCTCTCTCT CCTCTGAA TTAACCAAT GTAACTTCT AGAAGAAAG AACGTATAG TTCAACGGTT TCTTTAGCTT

FIGURE 7D

2401 TTTCTGGAAATTTAAAGTCGTG TGTTCACAGA GACCTGCGG GCAGGACGCT GCTTGTCCACC CACGGGAAG TGGTGAAGAT ATGTGACTTT GGAATTGGCTC
AARGACCTTA NATTCAGCAC ACAAGTGCT CTGGACCGGC GTCTCTTGA CCAACAGTGG GTGCCCTTC ACCACTTCTA TACACTGAAA CCTAAACGGAG

2501 GAGATATCAT GAGTGATTCC AACTATGTTG TCAGGGGCA TCCCGCTGTG CCTGTAAAT GGATGGCCCC CGAAGCCCTG TTTGAAGGCA TCACACACAT
CTCTATGTA CTCACTAAGG TTGATACAACTTTTCCGTT ACGGGCAGAC GGACATTTTA CCTACCGGG GCCTTCGGAC AACTTCCGT AGATGTGGTA

2601 TAAGAGTGAT-GTCTGGTCAT ATGGAATATT ACTGTGGAA ATCTTCTCAC TTGGTGTGA TCCTTACCCT GGCATTCCGG TTGATGCTAA CTTCACAAA
ATTCTCACTA CAGACCACTA TACCTTATTA TACACCCCTT TAAAGATG TACACACTT AGGATGGGA CCGTAAGGCC AACTACGATT GAAGATGTTT

2701 CTGATTCAAA ATGATTTAA AATGGATCAG CCATTTTATG CTACAGAGA AATATACATT ATATGCAAT CCTGCTGGGC TTTTGACTCA AGGAACGGC
GACTAAGTTT TACCTAAATT TTACCTAGTC GGTAAATAC GATGCTTCT TATATGTA TATTACGTTA GGACGACCGG AACTGAGT TCCCTTGGCG

2801 CATCCTTCCGTAATTGACT TCGTTTTAG GATGTAGCT GGCAGATGCA GAAGAAGGGA TGTATCAGAA TGTGGTGGC CGTGTTCGG AATGTCTCA
GTAGGAAGG ATTAACCTGA AGCAAAATC CTACAGTGA CCGTCTAGT CTCTCTGCT ACATAGCTT ACACCTACCG GCACAAAGCC TTACAGGAGT

2901 CACCTACCAA AACAGGCGAC CTTTCAGCAG AGAGATGGAT TTGGGGCTAC TCTCTCCCA GGCACAGGTC GAAGATTCGT AGAGGAACAA TTTAGTTTTA
GTGGATGGTT TTGTCGGCTG GAAAGTCCT TCTCTACCTA AACCCGATG AGAGAGGCGT CCGAGTCCAG CTTCTAGCA TCTCCTTGT AATCAAAAT

3001 AGGACTTCAAT CCGTCCACCT ATCCCTAACA GGCTGTAGAT TACCARAACA AGGTTAATTT CATCACTAAA AGAATATCTA TTATCAACTG CTGCTTCACC
TCCTGAAGTA GGGAGGTGGA TAGGGATTGT CCGACATCTA ATGGTTTTGT TCCAATTAAG GTAGTGATTT TCTTTTAGAT AATAGTTGAC GACGAAGTGG

3101 AGACTTTTCTG CTACAGAGCG
TCTGAAGAAGATATCTCTCC

29/35

FIGURE 8A

1 TCGGCTCCA CCGCCCAAG GAGATCAGA CTTGKKGCG CGAGGCCCG CCMACTCAG TTGGATCCT ACCGAGTGA GCGCGGCGCA TGGAGCTCC
 AGCCGAGGT GCGCGGTCC CTCTAGTCT GACCCCGCG GCTCCGCGG GATTGAGTC AGCCTAGGA TGGCTCACT CCGCGCGGT ACCTCAGGC
 M E L R

101 GGTGCTGCTC TGTGCGCTT GGTGCGCGC AGCTTTGGA GAGACCTGC TGAACACAA ATTGGAACCT GCTGATCTGA AGTGGTGAC ATTCCCTCAG
 CCAGGACGAG ACGACCGAA GCACCGCGC TCTGACCTT TCTGACGCG ACTTGTTT TACCTTGA CCACTAGCT TCACCCACTG TTAGGGAGTC
 5 V L L C W A S L A A A L E E T L L N T K L E T A D L K W V T F P Q

201 GTGAGCGGC AGTGGAGGA ACTGAGCGC CTGATGAGG AACAGCACAG CGTGCACCC TACGAAGCT GTGAGTGA GGTGCGCGG GGCAGGCGC
 CACTGCGCG TCACCTCCT TGAATCGCG GACTACTCC TTGCTGCTC GCACGCGTG ATGCTTACA CACTGCACT GGCAGCGGC CCGTCCGCG
 38 V D G Q W E E L S G L D E E Q H S V R T Y E V C D V Q R A P G Q A H

301 ACTGGCTTGG CACAGTTGG GTCCACGCG GGTGCGCGT CCAGTGTAC GCCAGCTGC GTTCACCAT GCTGAGTGC CTGTCCTTGC CTGCGGCTGC
 TGACCGAGC GTGTCACCC CAGGTACCG CCGCGCGCA GTGACATG CGGTGCGAG CGAGTGTGA CGAGTCAAG GACAGGAGC GAGCCGAGC
 72 W L R T G W V P R R G A V H V Y A T L R F T M L E C L S L P R A G

401 GCGCTCTGCG AAGGAGCCT TCACGTCTT CTACTATAG AGGATGCG ACACGCGAC GCGCTTACG CCAGCTTGA TGGAGACCC CTACATCAG
 CCGAGAGCG TTCTCTTGA AGTGGCAGAA GATGATCTC TCGTACGCC TGTCCGCGT CGCGAGTGC GTTGGACCT ACCTCTTGG GATGTAGTTC
 105 R S C K E T F T V F Y Y E S D A D T A T A L T P A W H E N P Y I K

501 GTGGACAGG TGGCGCGGA GCATCTACC CGAGCGCGC CTGGGCGGA GGCACCGCG AGGTGAATG TCAGAGGCT GGTCTTGGG CCGCTCAGCA
 CACTGTGCC ACCGCGCCT CGTAGAGTGG GCTTCGCG GACCCGCGT CCGGTGCGC TTCCACTTAC AGTCTTGA CCGAGACCT GCGAGTCTT
 138 V D T V A A E H L T R K R P G A E A T G K V N V K T L R L Q P L S K

601 AGGTGGCTT CTACCTGCG TTCCAGGACC AGGTGCGCT GATGCGCGT CTATCCCTGC ACCTCTTCTA CAAAGATGC GCCAGCTGA CTGTGACCT
 TCCGACCGA GATGACCGG AGGTCTTGG TCCACGAGC GTACCGGAC GATAGGAGC L S L H L F Y K K C A Q L T V N L
 172 A G F Y L A F Q D Q G A C H A L L S L H L F Y K K C A Q L T V N L

30/35

FIGURE 8B

801 TACTGCGGT AGGATGGCCA GTGGGCGGA CAGCGGTCA GGGCTGCA CTGTGCTCG:GGTTGAGG CAGGTAGGG GAAACACCAAG TGCCGAGCCT
 ATGAGGGCAC TCCTACCGGT CACCGGCTT CTGCGGCAGT GCGGACGTC GACAGGAGC CCCAGCTCC GTGACTCC CTGTGGTTC ACGGCTCGGA
 218 Y C R E D G Q W A E Q P V T G C S C A P G F E A A B G N T K C R A C

901 GTGCCAGGG CACTTCAAG CCCGTTCAG GAGAGGTG CTCGACCA TGCCAGCCA ATAGCCACTC TAAACCAAT GATCAGCG TCTGCCAGTG
 CACGGGTCCC GTGGAGTTC GGGGACGTC CTCTTCAG GACGGTGGT ACGGTGGT TATCGGTGAG ATTGTGTA CCTAGTGGC AGACGGTCAC
 272 A Q G T F K P L S G E G S C Q P C P A N S H S N T I G S A V C Q C

1001 CCGGTCGG TACTTCGG CAGGCAGA CCGCGGTGT GCACCTGCA CCACCCCTCC TTGGGCTCG CGGAGCGTG TTTCGGCT GAAAGGCTCC
 GTCGAGCCC ATGAGGGCCC GTGCTGTCT GTGGGCTCA CCGGGACCT GGTGGGAGG AGCCGAGG GCCTGCCACC AAAGGCGGA CTTCGCGAGG
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S R L N G S

1101 TCCTGACAC TGAATGGAG TGCCGCCCTG GAGTCTGGT GCCAGAGGA CTTCACTAC GCCTCCCT GCCTGAGTG CCGACCCGGA GGCTCTGTG
 AGGAGGTGG ACCTTACCTC ACGGGGAC CTCAGACCAC CGGCTCTCT GAGTGGAG CGGAGCGGA CGGCTCTAC GGCTGGGCT CCGAGGACAC
 338 S L H L E W S A P L F S G G R E D L T Y A L R C R E C R P G G S C A

1201 CGCCCTGGG GGGAGACCTG ACTTTGACC CCGGCCCGG GAGCTGGTG GAGCCCTGG TGTTGGTTG AGGCTACGT CCTGACTTCA CCTATACCTT
 GCGGAGGCC CCTCTGGAC TGAAATGG GCGCGGGC CTGGAGCAC CTCGGAGCC AGCACCAGC TCCCGATGCA GGACTGAAGT GGATATGGA
 372 P C G G D L T P D P G P R D L V E P W V V V R G L R P D F T Y T F

1301 TGAGTCACT GCATTGACG GGGTATCTC CTTAGCCAG GGGCCGTC CATTGAGCC TGTCATGTC ACCACTGACC GAGAGTACC TCCTGCACTG
 ACTCCAGTGA GGTAACTTC CCATAGCAG GATCGTGC CCGGCGAG CCGGAGGAG GTAACTCG ACAGTTACAG TGTTGACTGG CTCTCCATGG AGGAGTCAC
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

1401 TCTGACATCC GGTGACCG GTCTCACCC AGCACTTGA GCTGGGCTG GCTGGACCA GTGGGCTGT GCTGACTAC GAGTCAANT
 AGACTCTAGG CCNACTGCG CAGGATGGG TGCTGNACT CCGACCGGAC CCGACAGGG GCGGTGGGT CACCCGACA CGACTGATG CTCCAGTTTA
 438 S D I R V T R S S P S S L S L A W A V P R A P S G A V L D Y E V K Y

32/35

FIGURE 8D

2201 CCTCGGCTA AACGACGAC AGTTACAGT CATCCAGCTC GTGGCATGC TCGGGGCAT CGCTCGGC ATCGGTACC TTCCGAGAT GAGCTAGTC
 705 L R L N D G Q F T V I Q L V G M L R G I A S G M R Y L A E M S Y V

 2301 CACGAGACC TGGTGTCTG CACATCTTA GTCAACGCA ACCTGTCTG CAAAGTGTCT GACTTTGSC TTTCGGATT CTGGAGGAG AACTCTTCCG
 GTGGCTCTG ACCGACGAG GTGTAGAT CAGTTGTCT TGGACGAG GTTTCACAGA CTGAACCGG AAGGCTTA GAGCTCTCTC TTGAGNAGGC
 738 H R D L A A R N I L V N S N L V C K V S D F G L S R P L E E N S S D

 2401 ATCCACCTA CACGAGCTC CTGGAGGAA AGATTCCAT CCGATGGACT GCGCGGAG CGATTGCTT CCGAGTTC ACTTCGGCA GTGATGCTG
 TAGGTGGAT GTGTCTGAG GACCTCTCTT TCTAAGGTA GGCTACCTGA CGGGGCTCC GGTAAAGAA GGCCTTCAAG TGAAGCGGT CACTACGAC
 772 P T Y T S S L G G K I P I R W T A P E A I A F R K F T S A S D A W

 2501 GAGTTACGG ATTGTGATG GGGAGGTGAT GTCATTTTGG GAGAGGCTT ACTGGGACAT GAGCAATCAG GACGTGATCA ATGCCATTGA ACAGGACTAC
 CTCATGCCCC TAACACTACA CCTCCACTA CAGTAAACCC CTCTCGGCA TGACCTCTGA CTGTTAGTC CTGCACTAGT TACGTAACT TGCTCTGATG
 805 S Y G I V M W E V M S F G E R P Y W D H S N Q D V I N A I E Q D Y

 2601 CGGCTGCCCC CGCCCCAGA CTGTCCCACC TCCCTCCACC AGTCTATGCT GACTGTGTGG CAGAAAGACC GGATGCCC GGCCGCTTC CCCCAGGTGG
 GCGACGGG GCGGGGTCT GACAGGGTGG AGGGAGGTGG TCGAGTACGA CTCGACNACC GTCTTCTGG CTTAGCGGC CGGGCGAAG GGGTCCACC
 838 R L P P P P D C P T S L H Q L H L D C W Q K D R N A R P R F P Q V V

 2701 TCAGCGCCT GGACAGATG ATCGGAACC CGGCAGCCT CAATGCTG GCGCGGAGA ATGGCGGCG CTCACACCTT CTCTGGACC AGCGGAGCC
 AGTCGGGGA CTGTTCTAC TAGGCTTGG GCGGTGGA GTTTAGAC CGGGCTCTT TACCGCCCG GAGTGTGGA GAGGACCTGG TCGCCCTGG
 872 S A L D K M I R N P A S L K I V A R E N G G A S H P L L D Q R Q P

 2801 TTACTACTCA GCTTTTGGCT CTGTGGGGA GTGGCTTGG GCATCAAAA TGGGAGATA CGAAGAAAT TTCCAGCGG CTGGCTTTGG CTCCTTCGAG
 AGTCATGACT CGAAGACGA GACACCGCT CACCGAGCC CGGTAGTTT ACCTTCTAT GCTTCTTCA AAGGTGCGC GACCGAACC GAGGAGCTC
 905 H Y S A F G S V G E W L R A I K H G R Y E E S F A A A G F G S F E

FIGURE 8E

2901 CTGTGACGCC AGATCTCTGC TGAGGACCTG CTCGAAATCG GATCACTCTT GCGGACAC CAGAGAAA TCTTGGCCAG TGTCCAGCAC ATGAATGCC
 GACCAAGTCGG TCTAGAGACG ACTCTTGAC GAGGTTAGC CTCAGTGAGA CCGCCCTGTG GTCTCTTTT AGAACGGTC ACAGTGCTG TACTTCAGGG
 938 L V S Q I S A E D L L R I G V T L A G H Q K K I L A S V Q H M K S Q

3001 AGGCCAAGCC GGGAAACCCG GGTGGACAG GAGCACCGC CCGCAGTAC TGACCTGCA GNACTCCCA CCGCAGGAC ACCGCTCC CATTTCCGG
 TCCGTTCCG CCGTTGGGC CCACCTGTG CTCCTGCGG GCGGTGATG ACTGACGTG CTGAGGGT GCGTCCCTG TGGCGAGGG GTAAAGGCC
 972 A K P G T P G G T G G P A P Q Y O P A G T P H P R D T A S P F S G

3101 GCGAGAGTGG GCACTCACAG AGGCCCCAG CCGTGTGCC CCGTGGATTG CACTTGAGC CCGTGGGTG AGGAGTTGC ANTTGAGA GACAGATTT
 CCGTCTACC CCGAGTGTG TCGGGGTC GGGACACGG GGCACCTAC GTAACTCG GGCACCCAC TCTCAACG TTAACCTCT CTGTCTAA
 1005 A E W G L T E A P S P V P R W I A L O A R G V R S W Q F G E T G F

3201 GCGGTTCTG CCATATAGG AGGCAAT CACCCCCAG CCACCTCGG GNACTCAGA CCAAGGGTA GCGGCTTT CCGTCAGAC TGGTGTGAC
 CCGCCAGAC GGTATATCC TCCCTTTA GTGGGTG GGTGAGCCC CTGAGTCT GTTCCACT CCGCGGAA GCGAGTCTG ACCACACTG
 1038 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

3301 CAGAGGAAA GGAAGTCCC AACATCTCC AGCTCCCA GGTGCCCC TCACCTTAT GGTGGCTTC CCGCAGCCA AGAGAGTGT GACTCCCTG
 GTCTCCTTT CTTACAGGG TTGTAGAGG TCGAGGGT CCACGGGG AGTCGACTA CCGACGAG GCGTCTGT TTCTCACA CTGAGGAG
 1072 E E K E V P N I S Q P P Q V P P S P O W V R S R R P K R V O L P C

3401 CCAGTCCAG AGTGGGGG CTGTCCACG GGGCAAGAG GGTGTGAG GCCAGTGAC AATCATTT GGGTTGTAG TCCAACTTG CTGCTGTCAC
 GGTGAGTCT TCACCCCCC GACAGGGTCC CCGTTCTTC CCGACAGTC CCGTCACTG TTTAGTAC CCAACATC AGGTTGAC GACGACGTG
 1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C C H

3501 CACCAACTC ATCATTTTT TTCCCTTCTA ATGCCCTC CCGAGTGC TCGTTCATA TCGAGTTT TCGATTTT TTTTGGTCT TAATTTTCT
 GTGGTTGAG TTAGTAAATA AAGGAAACAT TTACGGAG GGGTGGAG ACGAGTAT AACTCCAA AACTCAAC AAAAACCA ATTAAATA
 1138 H Q T Q S F F S L V N A P P P A A A F I L K V F E F C F W S O F F S

34/35

FIGURE 8F

3601 CCCCGTCC TTTTGTTC TCGTTTGT TTTCTACCG TCCTGTGTC ACTTTGTGT TGGAGGGAAC CTGTTTCACT ATGGCTCTCT TTGCCCCAAGT
GGGGCAAGG AATAACAAAG AAGCAAAACA AAGAGATGC AGGAACAGTA TTCAACACACA ACCTCCCTTG GACAAAGTGA TACCGGAGGA ACGGGTTCA
1172 P F P F C F F V L F F Y R P C H N F V L E G T C P T H A S F A Q V

3701 TGAACAGGG GCCCATATC ATCTCTGTTT CCAGAACAGT GCCTTGTGCA TCCACATCC CCGACCCCG CCTGGACCC CCAAGCTGTG TCCTATGAAG
ACTTTCTCCC CGGTAGTAG TACAGACAAA GGTCTGTCA CGGACCACT AGGTGTAGG GGCCTGGGC GGACCTGGG GGTTCGACAC AGGATACTTC
1205 E T G A H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

3801 GCGTCTGGG TGAGGTAGTG AAGGGCGG TACTGTGTGG TGGACCCAG AACGGACGC CGGTGCTTGG AGGGTTCTT AAATTATATT TAAAAAGTA
CCACACCCC ACTCCATCAC TTTTCCCGC ATCAACCAAC ACCTTGGTC TTGCTGTGG GCCACGAACC TCCCCAAGAA TTAAATATAA ATTTTTCAT
1238 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

3901 ACTTTTCTA TAAATAMAG AAATGGAC GTGTCCACG TCAGGGGTA AAAAAAAA AAAAAAAA
TGAATAACAT ATTATTTT TTTTACCTG CACAGGGTG AGTCCCAT TTTTTTTT TTTTTTT
1272 F L Y K O K K H G R V P A P G V K K K K K

35/35

FIGURE 9 ARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPNRWTAPEAIQYRKFSAS

FIGURE 10 NVLVKSPNHVKITDFGLARLLEGEKEYNADGGKMPIKWMALCIIHYRKFTTHQS

FIGURE 11 NCMLAGDMTVCVADFGLSWKIYSGATIVRGCASKLPVKWLALGSLADNLYTVHS

FIGURE 12 NCLVGKNYTIKIADFGMSRNLVSGDY

FIGURE 13 TRNILVENENRVKIGDFGLTKVLPQDKEYYKVKEPGESPIFWYAPESLTFSLFSVSD

FIGURE 14 ARNILVNSNLVCKVSDFGMSRVLEDDPEAAVYTRGGKIPIRWTAPEAIYRKFTSASD

INTERNATIONAL SEARCH REPORT

PCT/US 93/00586

International Application No

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ⁶		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68, C12N15/11		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁷		
Classification System	Classification Symbols	
Int.Cl. 5	C12N ; C12Q ; C07K	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁸		
III. DOCUMENTS CONSIDERED TO BE RELEVANT⁹		
Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system' see the whole document --- -/--	1-7
<p>¹⁰ Special categories of cited documents : ¹⁰</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search		Date of Mailing of this International Search Report
07 JUNE 1993		02 -07- 1993
International Searching Authority		Signature of Authorized Officer
EUROPEAN PATENT OFFICE		ANDRES S.M.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p> <p>---</p>	1-7, 13-15
A	<p>GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRÄUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document</p> <p>---</p>	1-7, 13-15
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10</p> <p>---</p>	1-7
A	<p>MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document</p> <p>---</p>	1-7, 16-18
P,A	<p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document</p> <p>-----</p>	1-12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 93/00586

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
obscurities.
see additional page
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

OBSCURITIES

Claims 2,3,5 and 7 as filed in the application are in contradiction with the description. Indeed, they are dependent on claims 1,4 or 6 which are related to DNAs or proteins of human MEGAKARYOCYTIC origin, whereas they claim some sequences which are not from such an origin, if taking reference to page 8 of the description. Therefore the search of these claims has been made independently of claims 1,4 and 6.

Claims 14 and 15 relating to the nucleotide and the protein sequences of SAL-D4 are filed as being dependent on claim 10 which relates to SAL-S1. They have been interpreted and searched as depending on claim 13.

Claims 17 and 18 relating to the nucleotide and the protein sequences of LpTK 3 are filed as being dependent on claim 14 which relates to SAL-D4. They have been interpreted and searched as depending on claim 16.

Claims 20 and 21 relating to a DNA expression vector and a cell transformed therewith, are filed as being dependent on claim 17 which relates to a protein (LpTK 3). They have been searched as depending on claim 19.

Finally, claim 2, point a) relating to SAL-S1 refers erroneously to SEQ ID 7 which describes SAL-D4. It has been interpreted as referring on SEQ ID 5 and 17).

US 9300586
SA 69794

07/06/93

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9214748	03-09-92	EP-A- 0536350	14-04-93
